The Biometric Society

BIOMETRICS

Founded by the Biometrics Section of the American Statistical Association

Table of Contents

Number 3 September 1950 Volume 6

Material for *Biometrics* should be addressed to the Chairman of the Editorial Board, Institute of Statistics, North Carolina State College, Raleigh, N. C.; and material for Queries should go to "Queries", Statistical Laboratory, Iowa State College, Ames, Iowa, or to any member of the committee.

Articles to be considered for publication in *Biometrics* should be submitted in triplicate.

The Biometric Society is an international society devoted to the mathematical and statistical aspects of biology and welcomes to membership biologists, mathematicians, statisticians and others who are interested in its objectives. Through its six regional organizations the Society sponsors regional and local meetings. National secretaries serve the interests of members in Italy, Denmark and the Netherlands and there are many members "at large". Dues in the Society are as follows: Full membership including subscription to *Biometrics* is $4.50. Members of the Biometrics Section of the American Statistical Association who subscribe to the journal through that organization may become members of The Biometric Society on the payment of only $2.00 annual dues.

Annual subscription rates to non-members are as follows: For American Statistical Association Members, $3.00; for subscribers, non-members of either American Statistical Association or The Biometric Society, $4.00. Subscription and application should be sent to The Biometric Society, P. O Box 1106, New Haven 4, Connecticut, U.S.A.

Entered as second-class matter at the Post Office at New Haven, Connecticut. *Biometrics* is published quarterly—in March, June, September, and December—by The Biometric Society, Box 1106, New Haven 4, Connecticut, U.S.A.

THE CONSTRUCTION AND USES OF FRACTIONAL FACTORIAL DESIGNS IN INDUSTRIAL RESEARCH

O. L. Davies and W. A. Hay

Biological Laboratories, Imperial Chemical (Pharmaceuticals) Ltd., Blackley, Manchester

1. *Introduction*

An important application of statistical methods to industrial research is in the design and analysis of experiments in connection with the improvement of manufacturing processes. This applies to chemical and biological processes, formulation of pharmaceutical preparations and, in fact, to most types of industrial research. For example, in chemical research it is frequently required to determine the effect of certain changes in reaction conditions or methods of manufacture on the yield and quality of chemical products. Typical reaction conditions which might be varied are temperature of reaction, time of reaction, rates and methods of agitation, concentration and amounts of reactants, different catalysts, etc. Similar considerations arise in chemicals manufactured by fermentation processes, e.g. penicillin, streptomycin, industrial alcohol, lactic acid, etc., for which we may wish to examine the effect of changes in the conditions of fermentation on the yield and quality of the products. The object of all such investigations is usually to improve the yield or the quality of the product, or to produce the product more economically.

This type of research may involve the examination of many different factors and the problem is how best to design the experiments to examine the effects of these factors. In this type of investigation it is usually sufficient, in one experiment, to examine one change in each of the factors under investigation, for example, an increase *or* decrease in temperature of a fermentation process, an increase or decrease in the concentration and amounts of one or more of the various constituents of the medium, etc. The complete factorial design, where the factors are examined in all possible combinations, has long been

accepted as the best answer to this type of investigation provided the number of factors is not large. The number of combinations is doubled for every additional factor and for a large number of factors the number of combinations involved often renders a complete factorial design prohibitive and uneconomical. In the more expensive and time consuming processes even a 2^5 factorial experiment would usually be regarded as prohibitive. The main advantages of a factorial design are that (a) it contains a large degree of 'hidden replication'; for example in a 2^5 experiment each effect is estimated with the precision of an average of 16 comparisons of pairs of observations, and (b) it supplies estimates of all interactions. When the experimental error is large and/or when high order interactions are expected to be appreciable, then there is no satisfactory alternative to the complete factorial design. Frequently, however, experimental error is not large and the precision supplied by a complete factorial design is not required. Moreover, in many investigations high order interactions are not of appreciable magnitude, and in such cases the complete factorial design is unnecessary. The answer to experimental situations of this kind is in the use of 'Fractional Factorial Designs' or 'Fractional Replication'. Many papers have appeared during the past few years on this subject and some of these are listed in the references at the end of this paper. Unfortunately for the industrial research worker, these papers written largely from a mathematical standpoint, have tended to mask the essential simplicity of these very useful and valuable designs. The purpose of this paper is to draw attention to a simple method of constructing fractional factorial designs and to discuss their uses in industrial research.

2. *Construction of Fractional Factorial Designs.*

Even with complex industrial processes a design of eight observations can usually be carried out without any serious practical difficulty, and in general, such a design would be regarded as the minimum for satisfactory results. For three factors each at two levels, eight observations would constitute a complete factorial design and when only three factors require to be investigated, the complete factorial design would be carried out. Denoting the three factors in the usual way, by the capital letters A, B and C, the normal condition (or the lower level of each factor) by (1), and the change in condition of a factor (or the higher level) by the corresponding small letter, then the eight combinations of the complete factorial design are

$$(1),\quad a,\quad b,\quad c,\quad ab,\quad ac,\quad bc,\quad abc$$

consisting of (1), commonly called the 'control', a change in each factor

one at a time, two at a time and all three together. To calculate the effect of A, we simply subtract the mean of the four observations not containing 'a' from the mean of the four containing 'a', and similarly for the main effects of B and C. In other words each of these effects, apart from the factor $\frac{1}{4}$, is represented by the sum of four given observations minus the others. Also, all interactions are represented in a similar way and the effects can be conveniently set out in the following tabular form:—

TABLE 1.

Effect	Observations or treatment combinations							
	(1)	a	b	c	ab	ac	bc	abc
A	−	+	−	−	+	+	−	+
B	−	−	+	−	+	−	+	+
C	−	−	−	+	−	+	+	+
AB	+	−	−	+	+	−	−	+
AC	+	−	+	−	−	+	−	+
BC	+	+	−	−	−	−	+	+
ABC	−	+	+	+	−	−	−	+

The signs in each row are those to be attached to each observation for the purpose of calculating the given effects. For each main effect the signs for the treatments containing the corresponding small letters are plus and the remainder minus. The signs for each interaction are the products of the corresponding signs for the constituent main effects, eg. the signs for ABC are the products of the corresponding signs A, B and C. This table is the same, except in arrangement, as that given by Yates (ref. 3, p. 11). If the minus sign is used to denote the lower level or the normal condition of the factor, and the plus sign the higher level or the change in the condition of the factor, the first three lines of the table can also be used as a convenient form of representing the design. Thus for the first observation A, B and C are minus which represent the normal condition of all the factors; for the second observation, A is plus with B and C minus, therefore this denotes a change in A only, represented by 'a' in the former notation, and so on. The dual purpose of the table of signs is very convenient and, as will now be shown, may be used to facilitate the construction of fractional factorial designs.

Suppose it is required to examine four factors, A, B, C, D in eight observations, then this can be done by using the signs of one of the

interactions of Table 1, preferably ABC, to represent the effect of D. In other words, the (+) signs of the row ABC are made to represent the higher level of the fourth factor D and the (−) signs the lower level of D. This operation may be referred to as equating the factor D to the three-factor interaction ABC. The design then becomes:—

TABLE 2.

	(1)	*ad*	*bd*	*cd*	*ab*	*ac*	*bc*	*abcd*
A	−	+	−	−	+	+	−	+
B	−	−	+	−	+	−	+	+
C	−	−	−	+	−	+	+	+
D (= ABC)	−	+	+	+	−	−	−	+
AB (= CD)	+	−	−	+	+	−	−	+
AC (= BD)	+	−	+	−	−	+	−	+
BC (= AD)	+	+	−	−	−	−	+	+

This is one half of the complete factorial design for the four factors A, B, C, D. The method of constructing this is the same as that of confounding the interaction ABC between two blocks and identifying the blocks with the levels of the fourth factor D.

It is seen that the product of the signs of A and B is the same as the product of the signs of C and D, therefore, in equating ABC to D, AB has also been equated to CD, similarly AC has been equated to BD, and BC to AD. Moreover, the product of any three of A, B, C and D is equal to the fourth, and therefore A has been equated to BCD, B to ACD and C to ABD. (These are the familiar aliases discussed by Finney, 1, 2, and other authors). It is not possible from the above design to obtain an estimate of AB free of CD or *vice versa*, and similarly for any other such pair, and therefore, the above design should be used only when some of the interactions can be assumed negligible. This is not necessarily a limitation because in a given investigation the experimenter may know, from theoretical considerations or from previous experience with the same or similar processes, that certain interactions are not likely to be appreciable or are less likely to arise than others. For instance, in most chemical reactions, an interaction would be expected between the time of reaction and the temperature of reaction, particularly around those conditions which result in the highest yield or best quality of the product. Where the same comparison measures two effects it is usually possible to attribute

most if not all the comparison to one of the effects, the other being rejected from prior knowledge. With four factors under investigation there would usually be no difficulty in arranging the conditions so that any two-factor interactions likely to exist would be confined to interactions between three of the factors, that is to say, one of the factors would be free of interactions with the other three.

In Table 2, if A, B and C are the three factors amongst which some interactions may exist and D is the fourth factor having no interaction with the other three, then the design will estimate the four main effects and the interactions AB, AC and BC.

If it is required to examine five factors, then at least one other two-factor interaction, e.g. AB, must be assumed zero and the fifth factor E would then be equated to AB in the usual way, that is, the pluses of AB would correspond to the higher level of E and the minuses to the lower level. The resulting design is a quarter factorial design. If yet another interaction, say AC, may be assumed zero, a sixth factor may be included provided this additional factor does not interact with any of the other five, and in the limit, if all interactions are zero, seven factors may be examined by equating the additional factor to the remaining interaction BC.

Summarising the above, it is seen that the possibilities with eight observations are:—

(1) Seven factors if all interactions are negligible
(2) Six factors and one two-factor interaction if all other interactions are negligible
(3) Five factors and the interaction of one factor with each of two others if all other interactions are negligible
(4) Four factors and all two-factor interactions between any three of them if all other interactions are negligible
(5) Three factors and all the interactions between them.

3. *Statistical Significance of the Effects.*

In fractional factorial designs involving as few as eight observations, most of the degrees of freedom are used up in estimating the main effects and certain two-factor interactions, and there are very few, if any, degrees of freedom left to obtain an estimate of the experimental error. This is not a serious limitation, however, because there is in existence a vast amount of experience on chemical, physical and biological experimentation, both in the laboratory and on the plant, and, for the vast majority of processes, fairly reliable information already exists on the magnitude of the experimental error which may be used

to assess the significance of the effects. Where no prior information exists on the experimental error, then it would usually be necessary to carry out larger designs than of eight observations in order to obtain information on the experimental error.

4. *Aliases in a Fractional Factorial Design.*

The aliases in a design of eight observations representing half a factorial design have already been given. The same method could, if desired, be extended to a quarter and higher fractions of a factorial design but this tends to become rather tedious. If it is required to obtain a full list of the aliases then the best method is to derive the defining contrasts and use the method given by Finney (Refs. 1 and 2). These defining contrasts may be readily derived; for example, consider the design for five factors in eight observations obtained from Table 1 by equating D to ABC and E to BC, that is $D = ABC$ and $E = BC$. Multiplying the first by D and the second by E using the rules* of multiplication given by Finney (1, 2), we obtain

$$I = ABCD \qquad \text{and} \qquad I = BCE.$$

$ABCD$ and BCE are thus two of the defining contrasts and the third is their product ADE. The aliases of any effect may then be obtained in the usual way from the relation

$$I = ABCD = BCE = ADE$$

by multiplying this relation by the given effect (Finney, loc. cit.).

In the case of 7 factors in 8 observations, the identities are:—

$$D = ABC, \qquad E = BC, \qquad F = AC, \qquad G = AB$$

from which we see that $ABCD$, BCE, ACF and ABG are four of the defining contrasts and the remainder are obtained by multiplying these in all possible combinations generating 15 defining contrasts. The term 'alias' has a special meaning, and $D = ABC$ really means that D and ABC cannot be separated and the corresponding comparison can be used to estimate D only when ABC is negligible. More will be stated on this point later.

5. *Designs of Sixteen Observations.*

The method of constructing fractional factorial designs of eight observations may be readily applied to designs of 16 observations.

*The product of two or more effects is that given by the laws of ordinary algebra with the additional condition that $A^2 = B^2 = \cdots = 1$. Thus $(ABC)(CD) = ABD$.

Sixteen observations are sufficient for a complete factorial design for four factors each at two levels. Denote the factors by A, B, C and D then the complete factorial design is given by:

(1), *a*, *b*, *c*, *d*, *ab*, *ac*, *ad*, *bc*, *bd*, *cd*, *abc*, *abd*, *acd*, *bcd*, *abcd*.

The table of signs similar to Table 1, but in this case for sixteen observations, may be readily constructed, thus, for each main effect, combinations containing the corresponding small letters are plus and the remainder minuses and the signs for the interactions are obtained as before, by multiplying together the rows corresponding to the capital letters contained in the interactions. The table of signs is as follows:—

TABLE 3.

	(1)	*a*	*b*	*c*	*d*	*ab*	*ac*	*ad*	*bc*	*bd*	*cd*	*abc*	*abd*	*acd*	*bcd*	*abcd*
A	−	+	−	−	−	+	+	+	−	−	−	+	+	+	−	+
B	−	−	+	−	−	+	−	−	+	+	−	+	+	−	+	+
C	−	−	−	+	−	−	+	−	+	−	+	+	−	+	+	+
D	−	−	−	−	+	−	−	+	−	+	+	−	+	+	+	+
AB	+	−	−	+	+	+	−	−	−	−	+	+	+	−	−	+
AC	+	−	+	−	+	−	+	−	−	+	−	+	−	+	−	+
AD	+	−	+	+	−	−	−	+	+	−	−	−	+	+	−	+
BC	+	+	−	−	+	−	−	+	+	−	−	+	−	−	+	+
BD	+	+	−	+	−	−	+	−	−	+	−	−	+	−	+	+
CD	+	+	+	−	−	+	−	−	−	−	+	−	−	+	+	+
ABC	−	+	+	+	−	−	−	+	−	+	+	+	−	−	−	+
ABD	−	+	+	−	+	−	+	−	+	−	+	−	+	−	−	+
ACD	−	+	−	+	+	+	−	−	+	+	−	−	−	+	−	+
BCD	−	−	+	+	+	+	+	+	−	−	−	−	−	−	+	+
ABCD	+	−	−	−	−	+	+	+	+	+	+	−	−	−	−	+

An additional factor E may be introduced by equating it to the highest order interaction $ABCD$. This results in a half-factorial design with the defining relation $I = ABCDE$ obtained from $E = ABCD$ by multiplying both sides by E (Finney loc. cit.). From the way the design has been constructed it is seen that all main effects, all two and three-factor interactions between A, B, C and D may be estimated provided all interactions of E with A, B, C and D are negligible. From the defining relation we see that if all three-factor interactions are negligible, the design will estimate all main effects and the two-factor interactions between all five factors.

Additional factors may be introduced by equating them to the three-factor interactions least likely to be appreciable. When all the three

and four-factor interactions between A, B, C and D can be assumed negligible, as will usually be the case, there will be several interactions which can be used for additional factors. With one additional factor E, the obvious choice is to equate it to the four-factor interaction $ABCD$. When two additional factors E and F are introduced one of these could be equated to a four-factor interaction and the other to a three-factor interaction. A better design would however, result if the two additional factors are equated to three-factor interactions. This is seen from the following considerations. The first arrangement gives:

$$E = ABCD, \qquad F = BCD$$

from which the following defining relation is obtained:

$$I = ABCDE = BCDF = AEF \tag{1}$$

The second arrangement gives

$$E = BCD, \qquad F = ACD$$

with the defining relation:

$$I = BCDE = ACDF = ABEF \tag{2}$$

From the last term in (1) it is seen that

$$A = EF, \qquad E = AF \qquad \text{and} \qquad F = AE \tag{3}$$

whilst in (2) all main effects are clear of two-factor interactions. On the assumptions made that the interaction between E and F and of E and F with A, B, C and D are negligible, the first arrangement is quite satisfactory but the second arrangement gives a better separation of main effects and two-factor interactions. This illustrates the following rule:

If only one additional factor is to be introduced then equate this to the highest order interaction. If two or more factors are to be introduced then equate these first to the interactions of next to the highest order and when these have been used up then to the highest order interaction.

It is seen from Table 3 that as many as five additional factors may be introduced while still preserving all the two-factor interactions between A, B, C and D provided the additional factors do not interact amongst themselves and with A, B, C and D.

6. *Complete System of Fractional Factorial Designs.*

Any given fractional factorial design represents one of a system of such designs satisfying the same experimental requirements, and the

method described above supplies one of these, but the complete system may readily be obtained as will now be illustrated. In a design of 16 observations the fifth factor E was introduced by equating it to the four-factor interaction $ABCD$, that is $E = ABCD$. However, E could have been equated to $-ABCD$ and the resulting design would be similar except that all the signs in the row corresponding to $ABCD$ would be reversed. Both fractional factorial designs together constitute all the combinations of the complete factorial design of the five factors A, B, C, D and E, and the contrast between the two fractions represents the five-factor interaction $ABCDE$. When introducing two additional factors say $E = ABCD$ and $F = BCD$, the sign of either or both may be changed giving the following system of four fractional factorial designs

$$(4.1) \qquad E = ABCD, \qquad F = BCD$$

$$(4.2) \qquad E = -ABCD, \qquad F = BCD$$

$$(4.3) \qquad E = ABCD, \qquad F = -BCD$$

$$(4.4) \qquad E = -ABCD, \qquad F = -BCD$$

These four fractions together constitute the complete factorial design, and the contrasts between them represent the interactions $ABCDE$, $BCDF$ and their product AEF. When 'n' additional factors are introduced there will be 2^n suitable fractional factorial designs and the one to be used should be chosen at random. This can easily be achieved by tossing a coin to decide what sign to use for each additional factor.

The reverse process of dividing a complete factorial design between blocks such that the contrasts between the blocks correspond to a given set of interactions, is the problem of confounding used extensively in agricultural experimentation (Ref. Yates (3)). Parallel situations arise in industrial experimentation, for example, it is not unusual in the manufacture of chemicals to encounter long and short periodic trends in the yield and quality of products, with the result that comparisons between observations carried out within a short interval of time are less subject to error than comparisons between observations made at longer intervals. The accuracy of a long term experiment may be increased if all important comparisons are confined to comparisons between observations carried out within relatively short periods of time. In the case of a complete factorial experiment, this is achieved by dividing the experiment into blocks, each block occupying a relatively short period of time, in such a way that the comparisons between

the blocks correspond to unimportant interactions, and in this way the important comparisons are kept within blocks. The interactions corresponding to the comparison between the blocks are said to be confounded with blocks.

The two fractional factorial designs of 16 observations mentioned above and derived by equating E to $ABCD$ and $-ABCD$ respectively, represent the two blocks in the complete factorial design of the five factors A, B, C, D and E confounding the five-factor interaction $ABCDE$. Again, the system of four fractional factorial designs (4.1) to (4.4) represent the four blocks of a 2^6 factorial design confounding the interactions $ABCDE$, $BCDF$ and AEF.

Confounding is not confined to complete factorial designs and, if necessary, any of the spare degrees of freedom may be used to divide the fractional factorial design into smaller blocks. This is done in the usual way (Ref. Yates, (3) p. 18) and the method will be illustrated in the following example:—

7. *An Example.*

This example describes the design used in one of several investigations, on the plant scale, of the effect of various factors on the yield of penicillin. There are three main stages in the production of penicillin, namely, the production of the inoculum, the fermentation stage and the chemical extraction of the penicillin. The inoculum produced at the first stage is used in the second stage for the production of the penicillin. This investigation was concerned only with the first two stages of the process.

There were several factors which the biologists considered likely to have a beneficial effect on the efficiency of the process and the following five were chosen for this particular investigation:

Stage 1	*Preparation of Inoculum*
A	Concentration of Corn Steep Liquor
B	Amount of Glucose
C	Quality of Glucose
Stage 2	*Fermentation*
D	Concentration of Corn Steep Liquor.

There was insufficient corn steep liquor from one delivery for the whole design and it was decided to use two deliveries, one half of the design to be made from each delivery; it was therefore necessary to add another factor

E Quality of Corn Steep Liquor

There were four similar fermenters available for this experimental work and so another factor is added

F Fermenters 1, 2, 3 and 4.

No large differences were expected between either the fermenters or the deliveries of corn steep liquor and therefore no interaction was likely between these two factors and the remainder. This part of the investigation was confined to one change in each of the factors, the change being taken in the direction considered, by the biologist, to be most likely to improve the yield. The changes in any case were relatively small so that no large interactions were likely to arise. As a safeguard it was decided to retain as many as possible of the interactions between A, B, C and D.

A design of 16 batches was considered the very minimum size to be of any practical value owing to the relatively large variation known to arise in the biological process. This number is sufficient for a complete factorial design involving four factors. A, B, C and D are the more important factors and since these are the only factors between which interactions are likely to exist, we construct a basic factorial design with these factors. This would enable the following 15 effects to be estimated.

Main effects:	A, B, C, D
Two-factor interactions:	AB, AC, AD, BC, BD, CD
Three-factor interactions:	ABC, ABD, ACD, BCD
Four-factor interactions:	$ABCD$

Factor E, that is the two qualities of corn steep liquor, is introduced by equating it to the interaction $ABCD$. There are four fermenters and to introduce these it is necessary to use up three degrees of freedom. This is really the standard problem of confounding the fractional factorial design of the five factors A to E into four blocks, each block corresponding to one fermenter. Two of the degrees of freedom may be chosen at will but the third will be the product of the two chosen ones. The best arrangement is two three-factor interactions and their product, which, unfortunately, must be a two-factor interaction. BC is the least likely of the interactions and therefore we chose ABD, ACD and BC to represent the comparisons between the fermenters. The method of dividing the observations into four blocks has been given by Yates (Ref. 3), which is: consider the signs of any two of these interactions, say, ABD and ACD (see corresponding rows in Table 4), and allocate those observations for which $ABD = (+)$ and $ACD = (+)$

to one fermenter, $ABD = (+)$, $ACD = (-)$ to another fermenter, $ABD = (-)$ and $ACD = (+)$, to the third fermenter and $ABD = (-)$, $ACD = (-)$, to the fourth.

As mentioned earlier, it is not unusual in the manufacture of chemicals to encounter trends with time, often of unknown causes, and in order to eliminate as much of this trend as possible it is desirable to introduce another system of confounding between four periods of times. To do this we use up the remaining two three-factor interactions ABC, BCD and their product AD. The comparison between the four times least likely to be appreciable is $(t_1 - t_2) - (t_3 - t_4)$ and so this particular comparison is equated to the two-factor interaction AD. This is achieved by considering the signs of ABC and BCD for the purpose of allocating the observations into the four blocks, thus $ABC = +$ and $BCD = +$ will fall in time block 1, $ABC +$ and $BCD (-)$ in time block 2, etc. The resulting design is then:—

TABLE 4

	Batch Number															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A	−	+	−	−	−	+	+	+	−	−	−	+	+	+	−	+
B	−	−	+	−	−	+	−	−	+	+	−	+	+	−	+	+
C	−	−	−	+	−	−	+	−	+	−	+	+	−	+	+	+
D	−	−	−	−	+	−	−	+	−	+	+	−	+	+	+	+
E	+	−	−	−	−	+	+	+	+	+	+	−	−	−	−	+
Fermenter	4	1	2	3	1	3	2	4	1	3	2	4	2	3	4	1
Time block	4	2	1	1	3	3	3	1	4	2	2	2	4	4	3	1

This design may also be written in the following form:—

TABLE 5

Time sequence	Fermenter			
	1	2	3	4
1	*abcde*	*b*	*c*	*ade*
2	*a*	*cde*	*bde*	*abc*
3	*d*	*ace*	*abe*	*bcd*
4	*bce*	*abd*	*acd*	*e*

This is the half factorial design *ABCDE* confounded between four fermenters and again between four times. The batches in each time block are made at approximately the same time and so this design practically eliminates the effect of time trends.

8. *The Use of Fractional Factorial Designs in Sequence.*

An important feature of most industrial experiments is that the observations are made in sequence, either singly or in sets of a few at a time, so that the observations of one set become available before the next need be started. Moreover, the time interval between the sets of observations is usually short, often a matter of days or even hours. In this way industrial experiments differ markedly from agricultural field experiments for which a whole year must elapse between successive trials. For this reason agricultural field experiments must be self-contained resulting in rigid complex designs. Once a field experiment in agriculture has been started it is not usually possible to change or modify the design but in most industrial work a high degree of flexibility exists because the situation may be reviewed after every observation or set of observations comes to hand. It is not necessary to adhere strictly to the design drawn up at the outset of an experiment but the design may be modified as the result of information gained from the earlier observations.

Assume that a complete factorial design has been carried out and the change in one or more of the factors produces a large effect on the yield of a certain product. In such a case appreciable interactions are likely to arise between this factor and the others so that only one half of the design is of use for assessing the effects of the other factors. If the effect of the factor is a large reduction in yield of product, the loss in yield on the plant scale would represent a large financial loss. Since a factor with a large effect can usually be detected with relatively few observations, much of the experimenter's time will have been needlessly wasted. If, instead of a full design, only a half-factorial design had been carried out, the loss would have been reduced to half, and for a quarter replicate, the loss would have been still further reduce and so on. If, on the other hand, no single factor produces a large effect, nothing is lost by carrying out a fractional factorial design initially because, if this design does not give sufficient information, we can follow on with other fractional factorial designs of the same system until sufficient precision is obtained.

The procedure to be followed when investigating a number of factors, even when a full factorial design may ultimately be required,

is to divide the factorial design into blocks confounding higher order interactions and carry out these blocks in succession, examining each block after it is completed. If the first block gives all the information required then other blocks need not be carried out. Another possibility is that one or more of the factors may give a large effect in which case all further work would be confined to the more favourable levels of these factors. The experiment is then redesigned with fewer factors or, if required, with other factors to replace the ones dropped after the first set. Examples appear in a paper by Floyd (11).

9. *Separation of Aliases.*

One purpose of continuing the experiment with further fractional factorial designs in the same system may be to obtain greater precision in the estimates of the main effects and certain interactions, but the addition of a second fractional factorial design to the first also has the effect of separating the aliases. For instance, in a quarter factorial design, each degree of freedom represents a group of four effects, that is to say, each effect has three other effects as aliases. Now two quarter factorial designs together form a half factorial design for which each effect has one alias. The second fractional factorial design has thus produced a separation of the aliases and twice the number of effects can be independently estimated. There are usually several designs to choose from for the second set, and different pairs of fractional factorial designs in the same system produce different types of separation of the aliases. Consider for example a design of eight observations covering five factors A, B, C, D and E derived by equating D and E to the interactions ABC and BC respectively. There are four possible designs of this type as shown earlier (section 6) and these are as follows:—

(5.1) $\left.\begin{matrix} D = ABC \\ E = BC \end{matrix}\right\}$ Defining relation: $I = ABCD = BCE = ADE$

(5.2) $\left.\begin{matrix} D = -ABC \\ E = BC \end{matrix}\right\}$ " " $I = -ABCD = BCE = -ADE$

(5.3) $\left.\begin{matrix} D = ABC \\ E = -BC \end{matrix}\right\}$ " " $I = ABCD = -BCE = -ADE$

(5.4) $\left.\begin{matrix} D = -ABC \\ E = -BC \end{matrix}\right\}$ " " $I = -ABCD = -BCE = ADE$

The aliases for each of the above fractional factorial designs can be derived in the usual way (Finney, loc. cit.); for example, the aliases for the main effect A in the above designs are, respectively,

(6.1) $$A = BCD = ABCE = DE$$

(6.2) $$A = -BCD = ABCE = -DE$$

(6.3) $$A = BCD = -ABCE = -DE$$

(6.4) $$A = -BCD = -ABCE = DE$$

It is necessary to know precisely what is meant by an 'alias'; this term now in general use, may be misleading because $A = BCD$ does not mean that A is equal to BCD but that A and BCD cannot be estimated separately, and the corresponding degree of freedom does in fact measure $(A + BCD)$, that is, the sum of the two effects. In the above system of fractional factorial designs it is easy to see that the group of effects corresponding to the main effect A are respectively:—

(7.1) $$A + BCD + ABCE + DE \qquad (= x_1)$$

(7.2) $$A - BCD + ABCE - DE \qquad (= x_2)$$

(7.3) $$A + BCD - ABCE - DE \qquad (= x_3)$$

(7.4) $$A - BCD - ABCE + DE \qquad (= x_4)$$

Denote the magnitude of these effects by x_1 , x_2 , x_3 and x_4 respectively. If we combine (7.1) and (7.2) we see that the sum gives $2(A + ABCE)$ and the difference gives $2(BCD + DE)$. We have, therefore, separated A from the two-factor interaction DE, and A has a four-factor interaction as an alias. The combination (7.1) and (7.3) also separates the main effect A and the two-factor interaction DE, and gives $2(A + BCD)$, $2(ABCE + DE)$. This however, is not quite so good because A now has a three-factor interaction as an alias.

When a further fractional factorial design is carried out in addition to the first two, a still further separation of the aliases occurs. For example, assuming that designs (5.1), (5.2) and (5.3) have been carried out, then if the same effect group as previously is considered, it is seen that the three effects A, DE and BCD have been separated, each having the four-factor interaction $ABCE$ as its alias. Thus, from (7.1) and (7.2) we obtain $2(A + ABCE)$, from (7.1) and (7.3), $2(DE + ABCE)$, and from (7.2) and (7.3), $2(BCD - ABCE)$. It must be noted that only two thirds of the observations are used for these estimates and the conclusion is that if A, DE and BCD are to be separated, then this can only be done with an efficiency of $66\frac{2}{3}\%$. When the experimental error is not large, this loss of efficiency is more than offset by the information gained on the other interactions. If, however, the three-factor interactions in addition to the four-factor interactions can be

assumed zero and only estimates of A and DE are required from relations 7.1–7.4 of the $\frac{3}{4}$ factorial design, then putting $BCD = ABCE = 0$ the following expressions are obtained:—

$$A + DE = x_1 \tag{8.1}$$

$$A - DE \quad x_2 \tag{8.2}$$

$$A - DE = x_3 \tag{8.3}$$

Averaging the last two expressions gives:—

$$A - DE = (x_2 + x_3)/2$$

whence $$A = \tfrac{1}{2}x_1 + \tfrac{1}{4}x_2 + \tfrac{1}{4}x_3$$

and $$DE = \tfrac{1}{2}x_1 - \tfrac{1}{4}x_2 - \tfrac{1}{4}x_3$$

Since each x is a difference between two means of four observations each, the variance of these estimates is $\frac{1}{2}[(\frac{1}{2})^2 + (\frac{1}{4})^2 + (\frac{1}{4})^2]\sigma^2 = 3\sigma^2/16$. The smallest possible variance for an estimate of one effect from 24 observations is $\sigma^2(1/12 + 1/12) = \sigma^2/6$, therefore, the efficiency of the estimates of A and DE in the $(\frac{3}{4})$ factorial design assuming that all three and four-factor interactions are zero, is $88\frac{8}{9}\%$. When of course, all interactions other than those between A, B and C can be assumed negligible, then all main effects and the two-factor interactions between A, B and C can be estimated with 100% efficiency in the $(\frac{3}{4})$-factorial design. When the fourth set is carried out, the factorial design is complete and all effects can then be separated and estimated with 100% efficiency. The above considerations may be readily applied to all the other main effects B, C, D and E and the interactions AB and AC. For example the effect groups containing interactions AB are:—

$$AB + CD + ACE + BDE \tag{9.1}$$

$$AB - CD + ACE - BDE \tag{9.2}$$

$$AB + CD - ACE - BDE \tag{9.3}$$

$$AB - CD - ACE + BDE \tag{9.4}$$

The first two fractional factorial designs will separate the two-factor interactions with 100% efficiency. The first three will separate the two-factor interactions and one three-factor interaction, the other being assumed zero. This can be done with $66\frac{2}{3}\%$ efficiency. If all three-factor interactions are zero, the first three fractional factorial designs will estimate AB and CD with $88\frac{8}{9}\%$ efficiency.

Similar considerations may be applied to any sequence of fractional

factorial designs. This method gives effectively a complete and simple solution for all combinations of fractional factorial designs belonging to the same or different systems of confounding for the 2^n class of factorial designs (Refs. 6, 10).

Acknowledgments are due to Mr. G. E. P. Box with whom the authors have had frequent discussion on the use of fractional factorial designs.

REFERENCES

(1) Finney, D. J. The fractional replication of factorial arrangements. *Annals of Eugenics, 12, 4,* 291–301.
(2) Finney, D. J. Recent developments in the design of field experiments III. Fractional replication. *Journal of Agricultural Science, 36, 3,* 184–191.
(3) Yates, F. The design and analysis of factorial experiments. *Technical Communication No. 35 Imperial Bureau of Soil Science,* 1937.
(4) Plackett, R. L. and Burman, J. P. The design of optimum multifractorial experiments. *Biometrika,* XXXIII, IV., 305–325.
(5) Brownlee, K. A., Kelly, B. K., and Lorraine, P. K. Fractional replication arrangements for factorial experiments with factors at two levels. *Biometrika,* XXXV, III *and* IV., 268–282.
(6) Kempthorne, O. A simple approach to confounding and fractional replication in factorial experiments. *Biometrika,* XXXIV, III *and* IV., 255–272.
(7) Nair, K. R. and Rao, C. R. Confounded designs for asymmetrical factorial experiments. *Science and Culture, 7,* 312–313.
(8) Rao, C. R. Factorial experiments derivable from combinatorial arrangement of arrays. *Supplement to the Journal of the Royal Statistical Society,* IX, *1–2,* 128–139.
(9) Rao, C. R. On hypercubes of strength d and a system of confounding in factorial experiments. *Bulletin of the Calcutta Mathematical Society, 38,* 67–78.
(10) Banerjee, K. S. A note on weighing designs. *Annals of Mathematical Statistics,* XX, *2,* 300–304.
(11) Floyd, J. C. Penicillin formulations: The efficacy of oily injections. *Journal of Pharmacy & Pharmacology, 1, 11,* 747–756.

SOME REMARKS ON ANIMAL POPULATION DYNAMICS

P. A. P. Moran

Institute of Statistics, Oxford University

The present paper is not a survey of the vast field of investigation into the dynamics of population change but an attempt to discuss a few specific problems in the hope of suggesting some new lines of investigations to those better qualified than the author to carry them out. Any investigation into population dynamics must necessarily involve mathematics in some way or other, and it is convenient and perhaps illuminating to begin by classifying the various ways in which mathematics is applied to biological problems of this kind. Moreover, such a classification will be found to hold equally well for the applications of mathematics to economics.

Without begging any philosophical questions, we may distinguish between "a priori" and "a posteriori" methods. In a priori applications of mathematics to biology, an attempt is made to describe the main features of a biological system in terms of a mathematical model involving a set of equations. From this model, we then try to make deductions which can be verified by observation or experiment. Such deductions need not be numerical (if they are, they are usually very difficult of verification) but they may often be qualitative—e.g., ultimate extinction of a population, existence of oscillations, and so on. The difficulty and dangers of taking into account all the factors in a situation without making the mathematical model very complicated are obvious, yet the more complicated the model, the more difficult is it of verification.

Such models may be further classified according to whether they do or do not involve probability, and are then described as "probabilistic" and "deterministic" respectively. Examples of the latter are the equations of Volterra and others describing predator-prey relationships[1]

[1]This theory is described as deterministic because once the constants and the initial values of the population densities are given, the development of the situation is determinate. Probability theory may be used to determine the functional relationships in the model e.g. by the consideration of "random encounters", etc.

and the matrix theory of a single population given by P. H. Leslie (1945), (1948). An example of a "probabilistic" model is provided by the evolutive stochastic processes studied by Feller (1939), D. G. Kendall (1947), (1948a), and (1948b), and others.

Alternatively we may begin a study of animal populations from actual numerical data: we must then use statistical methods. Here again we may eschew probability, and using only those statistical methods which are of *descriptive* value, but this does not take us very far. In order to estimate properties of our populations and to assess the provision of these estimates, we must introduce probability ideas and use the inferential techniques of statistical science.

Leaving aside the large field of sampling methods for the estimation of population density (e.g., the theory of capture—re-capture methods) which is obviously of great methodological importance to ecologists, an example of this type of problem arises from the statistical analysis of population cycles. One of the best sets of records showing cyclic changes are those on the trapping of lynx in Canada for which figures are given by Elton and Nicholson (1942). Perhaps the most suitable procedure for the analysis of such data would be to suppose these figures generated by some kind of stochastic process. The constants involved in such a model could then be estimated. Elton and Nicholson give the numbers of lynx caught in various regions of Canada over a long interval of time. These numbers show remarkably stable oscillations. The logarithms of the numbers have oscillations which are reasonably symmetrical about their mean, and an autoregressive scheme could be fitted. Thus if u_t is the logarithm of the numbers of lynx caught in year t, in one of the regions considered, we might try to find out whether they would be reasonably fitted by a scheme of the type

$$u_t = au_{t-1} + bu_{t-2} + \epsilon_t \tag{1}$$

where a and b are constants and the ϵ_t are random terms, with zero means, which are serially uncorrelated. This is best done by fitting a and b by the method of least squares. The regression of u_t on more than the two previous terms could of course be considered. Equation (1) may then be used to predict what will happen in the future. The standard error of such a prediction can be calculated but increases as the epoch of the predicted value moves into the future. An unsolved statistical problem of great difficulty is to devise a satisfactory test for correlation between two such series. The ordinary test of significance for a correlation coefficient cannot be applied (Moran, 1949). These types of problem are also of great interest in econometrics, and much theoretical and empirical work remains to be done on them.

The dangers involved in constructing an a priori model of a biological situation are apparent, but we must not go to the opposite extreme and conclude that such methods are valueless. Not only may they suggest further lines of experimental investigation, but, if we give up hope of constructing valid quantitative theories, ecology will remain a collection of isolated facts.

Consider now the problem of explaining why cyclical behaviour does occur in animal populations. All or nearly all the mathematical theories constructed to account for these observations depend on the interaction between two or more populations of animals and undoubtedly this must play a major part in most observed cycles. In the present paper we shall consider that may happen in a single population independently of changes in other populations.

We suppose that the density of population is n, the number in some well defined region. When we neglect the probabilistic aspect of the problem there is no objection to regarding this as a continuous variable. Then the rate of increase of n is dn/dt, and, ignoring migration (or supposing its net effect to be zero), this must equal the number of births minus the number of deaths. These in turn are Bn and Dn where B and D are defined as the birth and death rates. We then have:

$$\frac{d(\log n)}{dt} = \frac{1}{n}\frac{dn}{dt} = B - D. \tag{2}$$

If B and D do not depend on n, $1/n\ dn/dt$ will be a constant (apart from random fluctuations in the environment), and the population will either increase indefinitely or extinguish itself.* Thus either B or D or both must be density dependent for the population to continue to exist indefinitely in some state of equilibrium (using that word, for the moment, in the loosest sense). Hence, factors controlling the level of a population must be density dependent. This statement will continue to hold even if we make (2) into a probabilistic model by (say) letting B and D be dependent on such random phenomena as weather. We therefore assume B and D to be single-valued functions of n, and write them $B(n)$ and $D(n)$. We then have:

$$\frac{dn}{dt} = n\{B(n) - D(n)\}. \tag{3}$$

We now assert that if $B(n)$ and $D(n)$ depend on n and *not on any other characteristics* of the population, oscillations cannot occur. For sup-

*This ignores the possibility that $B = D$, which will not happen in practice because of random fluctuations.

pose they did. Then at some times the population would be increasing and at others decreasing and it would be possible to find two times at which the densities were equal but in one of which it was increasing and one decreasing. From (3) this is impossible. To construct a theory which will explain cyclic behaviour we must modify (3). Before we discuss how this is to be done, we notice that the whole argument depends on the meaning of the terms involved and not on any empirical facts. To know what does in fact control population density and what is the mechanism of the production of cyclic behaviour will require much empirical investigation.

Now consider in what way it is possible to modify the model described by (3). We may suppose that $(B(n) - D(n))$ is dependent also on the density of some other population whose own rate of increase depends on n. This could occur if the other population was parasitic on the first, or in competition with it, and thus we get theories associated with the names of Volterra, Nicholson and Bailey and others. Alternatively, we may suppose that $(B(n) - D(n))$ depends not only on n but also on the age distribution. Models may be constructed mathematically, either by using integral equations (Lotka) or in terms of matrix theory (Leslie 1945). In these theories, if the birth and death rates are age specific but do not depend on n or the past history of the population, the only oscillations which can occur are soon damped out. To construct a theory which would account for self-sustaining oscillations in a single population, therefore, we must suppose that $(B(n) - D(n))$ is also dependent on the past history of the population. We may call this dependence "hysteresis".

As an example, consider a highly simplified model of a vole population. The role of hysteresis in this model will be considered later. The vole has a life of approximately one year. At the beginning of spring the population density is low. The numbers increase during the summer and decrease during the winter. If the number in a given area is measured at the same time every year, successive values generally increase for several years until there is decrease, known as the "crash", which usually occurs during one year only. The cycle then repeats, and has a period of the order of four years. Suppose that the densities at the beginning of spring are n_1 , n_2 , $\cdots$ these being the numbers in a given area. Suppose we assume a crudely deterministic model such as will result in the density n_{i+1} being some mathematical function of n_i alone (and thus not of the previous n's or of the previous history of the population up to the spring of year i). We write this function $n_{i+1} = f(n_i)$. It is natural to assume that as n_i increases from zero, n_{i+1} will also. But if $f(n_i) > n_i$ for all n_i , the population will con-

tinually increase and if $f(n_i) < n_i$ for all n_i , continually decrease. There must therefore be at least one point on the curve at which $n_i = f(n_i)$. We shall see later that, for self-sustaining oscillations to be set up, $f(x)$ must have its slope negative at this point and an increase of n_i must result in a decrease of n_{i+1} . We therefore assume that the

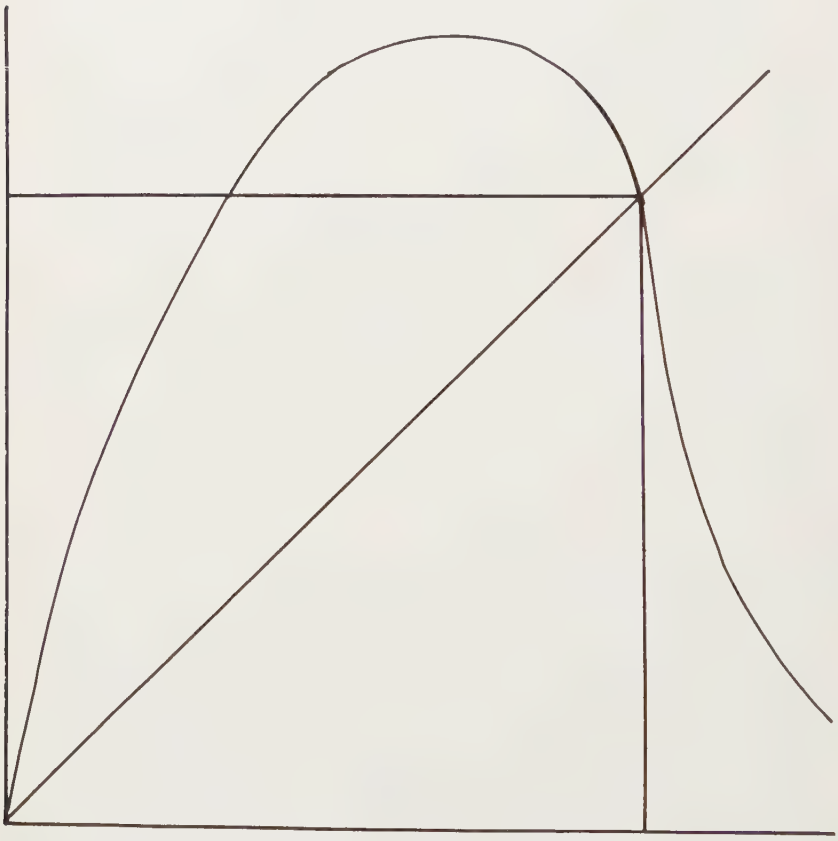

FIGURE 1

curve is of the form shown in Fig. I. and that, as x increases, $f(x)$ increases up to a maximum and then decreases continually without becoming zero. Let x_m be the value at which $f(x)$ has a maximum. With the possible exception of the initial value of the sequence, all the $n_i < f(x_m)$ and so $f(n)$ need not be considered for values of x greater than $f(x_m)$. Starting from any initial value n_1 we obtain successively

$$n_2 = f(n_1).$$

$$n_3 = f(n_2).$$

. . . .

The behaviour, oscillatory or otherwise, of the series n_1 , n_2 , $\cdots$ will be determined by n_1 and the function $f(x)$.

Let n_s be the value of n (assumed unique) for which $n = f(n)$. This is represented by the point where the curve $y = f(x)$ crosses the line $y = x$. If the initial value equals n_s , so will all subsequent values (ignoring the effects of random fluctuations and external effects). We may call such a process "stationary" because, although the population density rises and falls during the year, it does not change from year to year. As stated above, observed vole populations do not behave like this at all but show a more or less steady rise during several years followed by a sudden decrease known as the "crash". With rare exceptions, this crash occurs in one year only, and we do not have

$$n_{t+2} < n_{t+1} < n_t .$$

This phenomenon is also reproduced in the model, for

$$n_2 = f(n_1) < n_1$$

implies $n_1 > n_s$ and $n_2 < n_s$ and therefore $n_3 = f(n_2) > n_2$.

Next consider the stability of the stationary process where the population density is constant and equal to n_s . Suppose that initially there is a small disturbance so that $n_1 = n_s + \delta n_1$ where δn is small. Then

$$n_2 = f(n_s + \delta n_s)$$

$$= f(n_s) + \delta n_1 f'(n_s), \qquad \text{approximately}.$$

For stability we must have

$$\left| \frac{\delta n_2}{\delta n_1} \right| < 1$$

where $\delta n_s = n_2 - n_s = \delta n_1 f'(n_s)$
and we must therefore have $| f'(n_s) | < 1$.
If this is true, there will be a region around n_s such that if n_1 is in this region subsequent n's will converge to n_s . We then say that such a process is "stable".

Though $f'(n_s) < 1$ (because $y = f(x)$ crosses the line $y = x$ from

above), it is not impossible that $f'(n_s) < -1$. No stable population, stationary from year to year, is then possible, and the population must oscillate. No "optimum level" of population density can then be defined. The oscillations are bounded above and below, because $f(x)$ has a maximum $f(x_m)$ and, whatever all subsequent values (with the possible exception of a few immediately following n_1) must be between $f(f(x_m))$ and $f(x_m)$. The model thus looks like giving the kind of oscillations which are observed.

We may proceed further and consider oscillations which are definitely cyclic, as in a process for which there is a cycle of

values $$n_1 , n_2 , \cdots n_p , n_1 , \cdots$$

where $$n_1 < n_2 < \cdots < n_p > n_1 .$$

This process involves only a single crash (from n_p to n_1) in each cycle.

If we write $f_p(x) = f(\cdots f(f(x))\cdots)$ where the f occurs p times,

$$n_{p+1} = f_p(n_1).$$

The cycle can be shown to be stable if

$$\left| \frac{d}{dx} f_p(x) \right|_{x=n_1} < 1$$

which is equivalent to

$$| f'(n_1) \cdots f'(n_p) | < 1.$$

This condition for stability in a process of this kind occurring in in economic theory has been given by Leontief (Sammuelson (1947) p. 391). That stable cycles can exist with suitable chosen $f(x)$ can be easily shown mathematically. I have carried out numerical and mathematical investigations on a number of such functions which verify this conclusion, but, in the present state of knowledge of what actually occurs, a description of them would scarcely be of value to the ecologist. In principle the shape of $f(x)$ might be estimated from experimental observations, but considerable difficulties will arise. One of these is that both the observations (x, y) where y is a presumed value of x will probably be subject to experimental error. This will have the consequence that it will be difficult to establish the slope of $f(x)$ with any certainty.

We have not proved that there do not exist stable cyclic solutions in which two or more distinct crashes occur: thus we might imagine a process

$$n_1 , n_2 , n_3 , n_4 , n_1 , \cdots$$

where $$n_1 < n_2 > n_3 < n_4 > n_1$$

but n_1 , n_2 , n_3 and n_4 are all unequal. Criteria for the existence of such cycles in terms of $f(x)$ are not known.

This theory depends essentially on the assumption that a high density at the beginning of spring has such an adverse effect on the death and (or) birth rates that an unduly low density the following year must result. Thus we can choose n_1 , n_1' such that $n_1 < n_1'$ but $n_2 = f(n_1) > n_1' = f(n_2)$. If we follow the course of the two densities during the year, there must be a point where they are equal but their rates of change unequal. This implies that the birth and (or) death rates of the two populations, at this point, cannot be solely dependent on the density and that there must be some kind of historical dependence or hysteresis. That such a phenomenon is necessary to explain oscillations arising from a purely intra-specific origin has already been pointed out above. It has been discussed (mainly with reference to interspecific relationships) by Solomon (1949) under the name of "Suppression". Some empirical evidence that it can occur in laboratory insect populations can be deduced from the experiments of Pearl, Miner and Parker (1927).

Some inadequacies in the above theory must be pointed out. We have assumed that future changes in population density are uniquely determined by the density at the beginning of spring, and thus that there is no hysteresis effect across this point. This is probably untrue as, for example, the density during the previous winter may have some long term effect on the reproductive capacity of the survivors. If this is important, we must either find a function $f(x)$ in terms of some other point in the cycle across which the hysteresis is assumed to be negligible (for example the end of summer) or construct a more complicated model.

Another defect in the theory is the neglect of the age distribution. If the death and birth rates are not only age specific but also such that their age-specific dependence is itself dependent on the population density, it would not be difficult to construct a mathematical theory which would result in self-generated oscillations. In such a case, the "hysteresis" effect arises in part from the changes in the age distribution. In practice both these assumptions are probably true, but, in default of any method of estimating their importance, it does not seem at present worth while setting up a mathematical model which would be much more complicated than the above. In the meantime it is to be hoped that further experimental work will be done and that the above

mathematical discussion may direct attention to some of the more important factors. A mathematical scheme of the kind considered in the present paper may also be applicable when we consider successive generations instead of successive epochs of time. Some interesting experimental results along these lines have been described by Nicholson (1950).

I am much indebted for very helpful suggestions and criticisms from members of the Oxford University Bureau of Animal Population, who are not to be held responsible, however, for any opinions expressed above.

REFERENCES

Elton, C. and Nicholson, M. *Jour. Animal Ecology.* II, 215–244.
Feller, W. *Acta Biotheoritica. 5.* 11–40.
Kendall, D. G. *J. R. Statistical Society. 110.* 130–137.
Kendall, D. G. *Biometrika. 35.* 6–15.
Kendall, D. G. *Biometrika. 35.* 316–330.
Leslie, P. H. *Biometrika. 33.* 183–212.
Leslie, P. H. *Biometrika. 35.* 212–245.
Moran, P. A. P. *Jour. Animal Ecology. 18.* 115–116.
Nicholson, A. J. *Nature.* March 25. 476, 1950.
Pearl, R., Miner, J. R., and Parker, S. L. *The American Naturalist. 61.* 289–318.
Sammuelson, P. A. *Foundation of Economic Analysis.* Harvard University Press.
Solomon, M. *Jour. Animal Ecology. 18.* 1–35.

Corrigenda: "Fitting a straight line when both variables are subject to error" by M. S. Bartlett, *Biometrics*, 5, 207–212, 1949.

(i) At the top of p. 210, the last factor in the formula should read

$$\left\{\frac{2}{k}+\frac{4}{n-2k}\right\}^{-1} \qquad \text{instead of} \qquad \left\{\frac{2}{k}+\frac{4}{n-k}\right\}$$

(ii) In the formula for t on p. 211, the factor

$$\left\{\frac{2}{k}+\frac{4}{n-k}\right\}^{-\frac{1}{2}} \qquad \text{should read} \qquad \left\{\frac{2}{k}+\frac{4}{n-2k}\right\}^{-\frac{1}{2}}$$

The author is indebted to J. M. Cameron for pointing out these errors; the numerical example is, however, based on the correct formulae.

COMPOUND SYMMETRY TESTS IN THE MULTIVARIATE ANALYSIS OF MEDICAL EXPERIMENTS*

D. F. Votaw, Jr.

Yale University, New Haven, Conn.

A. W. Kimball and J. A. Rafferty

USAF School of Aviation Medicine, Randolph Field, Texas

Summary. This paper presents tests of several statistical hypotheses which assert that experimental quantities are " 'stable' with respect to time". More generally, the hypotheses assert that experimental quantities satisfy certain "symmetry" conditions (see sections 1, 2).

Two of the statistical tests can be used to test hypotheses about "row effects" in the Analysis of Variance. However, these hypotheses—unlike the conventional Analysis of Variance hypotheses—*do not require that the experimental quantities be uncorrelated nor that there be homogeneity of variances* (see section 3).

An experiment in connection with investigations of cold injury is described. Tests of compound symmetry are employed in the analysis of the results. Readers who prefer non-mathematical discussions will find this illustrative example most informative. (See section 4).

In the Appendix the hypotheses are stated in full generality and expressions for the corresponding sample criteria are given (see sections A.1, A.2). For various special cases, the means, variances and exact cumulative distributions of the criteria are given together with χ^2-distributions that approximate the exact distributions (see section A.3).

1. *Introduction.* Occasionally the medical experimenter needs to test statistically whether the experimental quantities are " 'stable' with

*This paper is an expansion of a 15-minute paper (see Abstract No. 69 in [6, p. 81]) presented December 27, 1948, at a meeting in Cleveland, Ohio, of the American Statistical Association.

respect to time". The notion of such "stability"* requires interpretation; and in this paper two interpretations will be considered. In each it is presupposed that the observations have a normal multivariate distribution; thus a condition of stability can be expressed merely in terms of the true means, standard deviations or coefficients of correlation associated with the multivariate distribution.

Suppose that for each member of a sample of experimental animals % CO_2 in blood is measured at each of two times, T_1 , T_2 , and hematocrit is measured at each of three times, T_1' , T_2' , T_3' . Let X_1 , X_2 be the two CO_2 measurements and X_3 , X_4 , X_5 be the three hematocrit measurements; also assume that $(X_1 , X_2 , X_3 , X_4 , X_5)$ has a 5-variate normal distribution. Let the true mean and standard deviation of X_i be m_i and σ_i , respectively, and the true coefficient of correlation between X_i and X_j be ρ_{ij} . The assertion that there is "stability with respect to time" could be interpreted as follows:**

$$m_1 = m_2 \ ; \quad m_3 = m_4 = m_5 \ ; \quad \sigma_1 = \sigma_2 \ ; \quad \sigma_3 = \sigma_4 = \sigma_5 \ ; \tag{1.1}$$
$$\rho_{13} = \rho_{14} = \rho_{15} = \rho_{23} = \rho_{24} = \rho_{25} \ ; \quad \text{and} \quad \rho_{34} = \rho_{35} = \rho_{45} \ .$$

What is stated in (1.1) can be expressed as follows: for the two % CO_2 measurements the true means are equal and the true standard deviations are equal; for the three hematocrit measurements the true means are equal, the true standard deviations are equal, and the intercorrelations (between distinct measurements) are equal; and the intercorrelations between % CO_2 measurements and hematocrit measurements are equal.

When the two sets of times of measurement are the same, there is another relevant interpretation which takes into account whether the measurements are simultaneous. For example, suppose that % CO_2 is measured at two times, T_1 , T_2 (as before), and hematocrit is measured at only two times, T_1' and T_2' , and $T_1' = T_1$ and $T_2' = T_2$; then "stability" could be taken to mean:†

$$m_1 = m_2 \ ; \quad m_3 = m_4 \ ; \quad \sigma_1 = \sigma_2 \ ; \quad \sigma_3 = \sigma_4 \ ; \tag{1.2}$$
$$\rho_{13} = \rho_{24} \ ; \quad \text{and} \quad \rho_{14} = \rho_{23} \ .‡$$

*The hypotheses to be considered will first be discussed in terms of "stability". An alternative illustration will then be given (see (1.3)).

**See (2.1) for a restatement of (1.1).

†See (2.2) for a restatement of (1.2).

‡If simultaneity is taken into account and the sets of times of measurement are not identical but have at least one time common to two or more sets, then the situation is obviously a "mixture" of (1.1) and (1.2); however, such "mixed" cases will not be considered in this paper.

What is stated in (1.2) can be expressed as follows: for the two % CO_2 measurements the true means are equal and the true standard deviations are equal; for the two hematocrit measurements the true means are equal and the true standard deviations are equal; the intercorrelations between % CO_2 and hematocrit measured simultaneously are equal; and the intercorrelations between % CO_2 and hematocrit measured at distinct times are equal.

An illustration of the hypotheses will now be given in a situation in which the experimental quantities do not explicitly involve time. Suppose the anterior and posterior muscles of both hind legs of rabbits are to be weighed (see section 4). Let X_1 , X_2 , X_3 and X_4 be the anterior left, anterior right, posterior left and posterior right hind leg muscle weights, respectively. It is assumed that (X_1 , X_2 , X_3 , X_4) has a 4-variate normal distribution. In comparing the right and left legs, the following hypotheses might be of interest.

$$m_1 = m_2 \; ; \quad m_3 = m_4 \; ; \quad \sigma_1 = \sigma_2 \; ; \quad \sigma_3 = \sigma_4 \; ; \quad \text{and} \qquad (1.3)$$
$$\rho_{13} = \rho_{14} = \rho_{23} = \rho_{24} \; .$$

(1.3) asserts that: (a) the anterior muscle weights have the same means and the same variances; (b) the posterior muscle weights have the same means and the same variances; (c) the intercorrelations between anterior and posterior muscle weights are all equal. The experimenter might feel, however, that there are too many conditions in (c) and prefer the following hypothesis:

$$m_1 = m_2 \; ; \quad m_3 = m_4 \; ; \quad \sigma_1 = \sigma_2 \; ; \quad \sigma_3 = \sigma_4 \; ; \qquad (1.4)$$
$$\rho_{13} = \rho_{24} \; ; \qquad \text{and} \qquad \rho_{14} = \rho_{23} \; .$$

(1.4) asserts (a) and (b) above and the following two conditions: (1) the correlation of the two left-leg muscle weights equals the correlation of the two right-leg muscle weights; (2) the (left-leg anterior)-(right-leg posterior) correlation equals the (right-leg anterior)-(left-leg posterior) correlation.

The purposes of this paper are: (i) to present, in a simple form, sample criteria for accepting or rejecting hypotheses of the sort in (1.1), (1.2), (1.3), and (1.4), and for certain generalizations of such hypotheses (see A.1 in the Appendix); (ii) to give means, variances and distributions of the sample criteria (when the corresponding hypotheses are true) for various special cases; and (iii) to indicate the relation between two of the hypotheses, denoted by $H_1(m)$ and $\bar{H}_1(m)$, and certain hypotheses in the Analysis of Variance (see section 3).

Neither time nor a biological system need be involved in a situation to which the hypotheses are relevant; e.g., in Psychometrics a hypothesis of the sort stated in (1.1) can be used to test statistically whether three different forms of a psychological examination are interchangeable and have equal validities regarding both of two criterion measures (see [2, pp. 447–448]).

2. *Restatements of Hypotheses.* Inspection of (1.1) and (1.2) shows that in either interpretation of "stability" not only the number of quantities to be measured but also their *grouping* is important. Similar remarks apply to (1.3) and (1.4). Five quantities, $X_1 , \cdots , X_5$, are involved in (1.1) and they fall into two mutually exclusive groups: (X_1 , X_2), (X_3 , X_4 , X_5). In (1.2), (1.3), and (1.4) four quantities, X_1 , X_2 , X_3 , and X_4 , are involved, falling into two mutually exclusive groups: (X_1 , X_2), (X_3 , X_4). Convenient notations for these two groupings are: (2,3) and (2,2), respectively. (See A.1 in the Appendix for a full discussion of grouping). For the case in which there is only one group see [1].

Let $H_1(mvc)$ be the hypothesis stated in (1.1), (1.3), or any of the generalizations of (1.1) or (1.3) (see A.1 in the Appendix). Let $A^{ij} = A^{ji} = \sigma_i\sigma_j\rho_{ij}$; $\| A^{ij} \|$ is the variance-covariance matrix of the chance quantities; for example in (2.1), below, $A^{23} = \sigma_2\sigma_3\rho_{23}$, $A^{33} = \sigma_3^2$. For the grouping (2,3) $H_1(mvc)$ may be stated as follows:

$$m_1 = m_2 \; ; \quad m_3 = m_4 = m_5 \; ; \qquad \text{and}$$

$$\left\| \begin{matrix} A^{11} & A^{12} & \cdots & A^{15} \\ A^{21} & A^{22} & \cdots & A^{25} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ A^{51} & A^{52} & \cdots & A^{55} \end{matrix} \right\| = \left\| \begin{matrix} B & C & D & D & D \\ C & B & D & D & D \\ D & D & E & F & F \\ D & D & F & E & F \\ D & D & F & F & E \end{matrix} \right\| \qquad (2.1)$$

(2.1) is simply a restatement of (1.1). The broken lines in (2.1) indicate the grouping of the rows and columns in the variance-covariance matrix.

In certain cases we might wish to ignore the conditions on the means and assert merely that $\| A^{ij} \|$ has the form given in (2.1)*. This less restrictive hypothesis is denoted by $H_1(vc)$. In some situa-

*or in a generalization of (2.1) (see A.1).

tions we might take $H_1(vc)$ as given and merely assert that the means satisfy the conditions given in (2.1); this third hypothesis is denoted by $H_1(m)$.

Let $H_1(mvc)$ represent the hypothesis in (1.2), (1.4), or any of the generalizations of (1.2) or (1.4) (see A.1). $H_1(mvc)$ may be stated as follows:

$$m_1 = m_2 \; ; \quad m_3 = m_4 \; ; \qquad \text{and}$$

$$\left\| \begin{matrix} A^{11} & A^{12} & A^{13} & A^{14} \\ A^{21} & A^{22} & A^{23} & A^{24} \\ A^{31} & A^{32} & A^{33} & A^{34} \\ A^{41} & A^{42} & A^{43} & A^{44} \end{matrix} \right\| = \left\| \begin{matrix} G & J & T & U \\ J & G & U & T \\ T & U & K & L \\ U & T & L & K \end{matrix} \right\|. \tag{2.2}$$

We shall also consider hypotheses $\overline{H}_1(vc)$ and $\overline{H}_1(m)$, which have to $\overline{H}_1(mvc)$ the same relations, respectively, that $H_1(vc)$ and $H_1(m)$ have to $H_1(mvc)$.

For the grouping (2,2) $H_1(mvc)$ may be stated as follows:

$$m_1 = m_2 \; ; \quad m_3 = m_4 \; , \qquad \text{and}$$

$$\left\| \begin{matrix} A^{11} & A^{12} & A^{13} & A^{14} \\ A^{21} & A^{22} & A^{23} & A^{24} \\ A^{31} & A^{32} & A^{33} & A^{34} \\ A^{41} & A^{42} & A^{43} & A^{44} \end{matrix} \right\| = \left\| \begin{matrix} R & S & W & W \\ S & R & W & W \\ W & W & T & U \\ W & W & U & T \end{matrix} \right\|. \tag{2.3}$$

(2.3) is simply a restatement of (1.3). $H_1(vc)$ and $H_1(m)$ can be formulated easily for the (2,2) grouping by means of (2.3), e.g., $H_1(vc)$ asserts merely that the variance-covariance matrix has the form given in (2.3).

A normal multivariate distribution for which one or more of the six hypotheses $H_1(mvc)$, $H_1(vc)$, $\cdots$, $\overline{H}_1(m)$ is true, is said to have "compound symmetry". The symmetry is of Type I or Type II according as an H or an $\overline{H}$ hypothesis is applicable.* The sampling theory of

*The hypothesis of Type II symmetry was formulated on the basis of a description by J. Allan Rafferty, M.D. (USAF School of Aviation Medicine, Randolph Field, Texas) of the need in medical research for testing statistically whether a biological system is "stable".

criteria for hypotheses of compound symmetry is given in [2]. The sampling theory of criteria relevant to the case where there is only one group of variates is given in [1]; in this case the two types of symmetry are identical.

Sample criteria for the hypotheses $H_1(mvc)$, $H_1(vc)$, $\cdots$, $\overline{H}_1(m)$, are represented by $L_1(mvc)$, $L_1(vc)$, $\cdots$, $\overline{L}_1(m)$, respectively.* In section 4, $L_1(mvc)$, $L_1(vc)$, and $L_1(m)$ are given explicitly for the grouping (2,2) as functions of the sample means, sums of squares and sums of cross products.

3. *$H_1(m)$, $\overline{H}_1(m)$ and a Hypothesis in Analysis of Variance.* A sample 0_N $(X_{1\alpha}, X_{2\alpha}, \cdots, X_{t\alpha})$ $(\alpha = 1, \cdots, N)$ of size N can be exhibited in a txN array as follows:

$$\begin{array}{cccc} X_{11}, & X_{12}, & \cdots, & X_{1N} \\ X_{21}, & X_{22}, & \cdots, & X_{2N} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ X_{t1}, & X_{t2}, & \cdots, & X_{tN} \end{array} \qquad (3.1)$$

Each column, $(X_{1\alpha}, X_{2\alpha}, \cdots, X_{t\alpha})$, in (3.1) will be considered as a normal t-variate chance quantity, where the t variates may be correlated $(\alpha = 1, \cdots, N)$; any two X's in different columns will be assumed to be uncorrelated. In 3A, 3B, 3C, below, tests of hypotheses about "row effects" will be considered. 3A contains merely a discussion of a "row effect" test when the observations are uncorrelated; 3B deals with the case where for all columns there is a common correlation, ρ, among elements within a column; 3C deals with a generalization of 3B. The hypothesis treated in 3B is similar to H_m in [1]; the generalization in 3C is based on hypotheses similar to $H_1(m)$ and $\overline{H}_1(m)$.

3A. *Analysis of Variance Test for Equality of Row Effects.*

In one of the approaches to the Analysis of Variance (see [5, pp. 177–179]) it is assumed that the distribution of $X_{i\alpha}$ is normal with true mean $u + r_i + c_\alpha$ and variance σ^2 $(i = 1, \cdots, t, \alpha = 1, \cdots, N)$,

*The notation is the same as that in [2], wherein the subscript 1 in H_1 indicates that only one population is under consideration.

and that the X's are uncorrelated. The r's and c's represent row and column effects, respectively, and u is an effect common to all rows and columns. In such cases, the usual null hypothesis tested is that the row effects are equal, i.e., $r_1 = r_2 = \cdots = r_t$. For this test, the appropriate statistic is the variance ratio, F. (See (c) p. 179, in [5]).

3B. *An F-test for Equality of "Row Effects" When There Are Correlations.*

A test for equality of row effects can be made even when there are correlations among some of the X's in (3.1). In fact, the same F-test referred to above is still valid, under the following conditions: the true mean of $X_{i\alpha}$ is $u + r_i + c_\alpha$ $(i = 1, \cdots, t; \alpha = 1, \cdots, N)$; the true variance of $X_{i\alpha}$ is σ^2 for all i and α; the true correlation between $X_{i\alpha}$ and $X_{j\alpha}$ is ρ for all i, j, α; and $X_{i\alpha}$ and $X_{j\alpha'}$ are uncorrelated $(i, j = 1, \cdots, t; \alpha, \alpha' = 1, \cdots, N; \alpha \neq \alpha')$; $(X_{1\alpha}, \cdots, X_{t\alpha})$ has a normal t-variate distribution $(\alpha = 1, \cdots, N)$. The null hypothesis, say H'_m, is that $r_1 = r_2 = \cdots = r_t$. H'_m does not presuppose that the column effects are equal; however, if they are equal, then H'_m reduces to H_m in [1]. The statistic L_m, which equals $(N - 1)/(N - 1 + F)$ where F is the variance ratio (see [1, pp. 262, 265]), is a test of H_m. When H'_m is true, the distribution of L_m (also F) can be shown to be independent of the column effects; therefore L_m *is a test of* H'_m.

3C. *Test for Equality of Within-Group Row Effects When There Are Correlations.*

Assume there is a grouping* $(n_1, \cdots, n_h)$ of the t rows $(t = n_1 + \cdots + n_h)$. With the a-th group $(a = 1, \cdots, h)$ we associate a "group row effect", R_a; within the a-th group there are, say, n_a rows, which may be represented by 1, 2, $\cdots$, n_a —and with row z_a we associate a "within-group row effect", r_{az_a}. The statistic $L_1(m)$ (see section 4 and A.3) can be used to test for equality of within-group row effects for each group—i.e., to test that for every a, $r_{a1} = r_{a2} = \cdots = r_{an_a}$. Use of this test presupposes that: (A) the true mean of $X_{i\alpha}$ is $u + R_a + r_{az_a} + c_\alpha$ $(i = 1, \cdots, t; \alpha = 1, \cdots, N; a = 1, \cdots, h; z_a = 1, \cdots, n_a)$; and (B) for every α $(X_{1\alpha}, \cdots, X_{t\alpha})$ has a t-variate normal distribution with a common variance-covariance matrix having the pattern of symmetry in (A.1). The null hypothesis is that for every a, $r_{a1} = r_{a2} = \cdots = r_{an_a}$. When (A), (B), and the null hypothesis hold, it can be proved (by methods in [2]) that the distribution of $L_1(m)$ is independent of the column effects $c_1, \cdots, c_N$ and is therefore the same as the distribution of $L_1(m)$ given in [2]—moreover, given that (A) and (B) hold and $c_1 = c_2 = \cdots = c_N$, the hypothesis of

*(See A.1 for a discussion of grouping; also see the footnote after A.4).

equality of within-group row effects and the hypothesis $H_1(m)$ are identical.

The statistic $\overline{L}_1(m)$ (see A.2) can be used to test for equality of within-group row effects when the grouping is $(n^h)^*$ $(t = nh)$. The assumptions on which the test is based are: (A) (as above); and (B′) for every $\alpha(X_{1\alpha}, \cdots, X_{t\alpha})$ has a t-variate normal distribution with a common variance-covariance matrix having the pattern of symmetry described in the last paragraph of section A.1. The null hypothesis is the same as in the preceding paragraph. When (A), (B′), and the null hypothesis hold, the distribution of $\overline{L}_1(m)$ is independent of c_1, c_2, $\cdots$, c_N and is the same as the distribution of $\overline{L}_1(m)$ given in [2]. Also, given that (A) and (B′) are true and $c_1 = c_2 = \cdots = c_N$, the hypothesis of equality of within-group row effects and the hypothesis $\overline{H}_1(m)$ are identical.

4. *Application to a Medical Experiment.* Considerable attention has been focussed recently on the use of heparin and other anticoagulants in frostbite therapy. Particular interest has been manifested by the military services because of the expansion in the concept of global warfare to include frigid zones as possible combat areas. Along this line, the Department of Pathology at the USAF School of Aviation Medicine has been conducting a series of frostbite experiments with rabbits. In these experiments hind legs of rabbits are immersed in supercooled solutions and partially frozen. It has been determined that one of the most reliable measures of atrophy or necrosis is the change in weight of the muscle tissue. Unfortunately on a single rabbit it is not possible to obtain from the same leg a measurement of the initial muscle weight and a measurement of the muscle weight after exposure to cold. As a workable procedure experimenters have proposed that only one hind leg of each rabbit be frozen and that the weight of the corresponding muscle in the leg which is not frozen be taken as the initial weight of the muscle in the leg which is frozen. Although the method had been used with apparent success in previous work, the experimenters were anxious to test statistically the underlying hypotheses. The resulting experiment provides an excellent example of the use of compound symmetry tests in medical research.

Sixteen normal rabbits were selected at random. The rabbits were not subjected to any treatment, but the anterior and posterior muscles of both hind legs of each rabbit were removed and weighed. These measurements are recorded in Table 1. Let X_1, X_2, X_3, and X_4 be the anterior left, anterior right, posterior left and posterior right muscle

*See A.1.

weights, respectively. It is assumed that (X_1 , X_2 , X_3 , X_4) has a 4-variate normal distribution. Then the data in Table 1 constitute a random sample of size 16 from a 4-variate normal distribution.

Clearly, one of the hypotheses in which the experimenter is interested is that the two anterior muscle means are equal and the two posterior muscle means are equal.* In section 3 this hypothesis is designated $H_1(m)$ and may be written $m_1 = m_2 ; m_3 = m_4$*. A second hypothesis in which the experimenter is interested concerns the form of the variance-covariance matrix associated with (X_1 , X_2 , X_3 , X_4). He requires

TABLE 1
MUSCLE WEIGHTS (GRAMS)

Rabbit Number	Anterior		Posterior	
	Left Leg	Right Leg	Left Leg	Right Leg
1	5.0	4.9	15.0	15.2
2	4.8	5.0	14.2	14.3
3	4.3	4.3	12.8	12.8
4	5.1	5.3	14.4	14.6
5	4.1	4.1	11.0	11.0
6	4.0	4.0	12.5	12.6
7	7.1	6.9	19.6	19.5
8	5.9	6.3	15.9	15.8
9	5.3	5.2	14.1	13.8
10	5.3	5.5	14.5	14.8
11	5.3	5.5	16.3	15.7
12	5.9	5.9	16.4	16.2
13	6.5	6.8	18.6	19.0
14	6.3	6.3	18.1	17.4
15	6.6	6.6	17.3	17.5
16	6.2	6.3	18.1	17.7

that the variances of the two anterior measurements be equal and that the variances of the two posterior measurements be equal, but does not require that the anterior and posterior variances be equal. Furthermore, in this example it is reasonable to require that the intercorrelations between anterior and posterior muscle weights be all equal. This hypothesis concerning the variance-covariance matrix is designated $H_1(vc)$ and may be written

$$\sigma_1 = \sigma_2 ; \qquad \sigma_3 = \sigma_4$$

$$\rho_{13} = \rho_{14} = \rho_{23} = \rho_{24} .$$

*Given that the variance-covariance matrix satisfies certain symmetry conditions.

In the compound symmetry analysis the test for $H_1(m)$ assumes that $H_1(vc)$ is true. On the other hand the test for $H_1(vc)$ is valid whether or not $H_1(m)$ is satisfied. A third hypothesis $H_1(mvc)$ encompasses all the statements made in $H_1(m)$ and $H_1(vc)$, and a separate test of it may be made. The distinction to be made here is that $H_1(mvc)$ is a joint test of $H_1(m)$ and $H_1(vc)$ and should be rejected whenever either or both are not satisfied. A test of $H_1(m)$ alone would provide the experimenter with some knowledge of the interchangeability of the right and left leg muscle weights; but even if $H_1(m)$ is true, he may hesitate to use the right and left legs interchangeably if $H_1(vc)$ is not satisfied. In such cases it is probably best to make a final judgment only after all three hypotheses have been tested. The computational procedures which are given below are similar to those which would be employed in other problems with different groupings of the variates or for testing other hypotheses [$\overline{H}_1(m)$, $\overline{H}_1(vc)$, $\overline{H}_1(mvc)$].

The computations are described and illustrated in stepwise fashion, and a slightly different procedure is required for each of the three hypotheses. We have (a) a set of variates (X_1 , X_2 , X_3 , X_4), (b) a grouping of the variates into two groups of two each, (c) a significance level, say $\alpha(0 < \alpha < 1)$, and (*d*) a sample, say

$$\begin{array}{ccc} X_{11} , & \cdots , & X_{41} \\ X_{12} , & \cdots , & X_{42} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ X_{1N} & \cdots , & X_{4N} \end{array}$$

of N values of $(X_1 , \cdots , X_4)$. For testing $H_1(vc)$ the procedure is as follows:

(1) Compute $v_{ij} = v_{ji} = \sum_{\alpha=1}^{N} (X_{i\alpha} - \overline{X}_i)(X_{j\alpha} - \overline{X}_j)$,

$(i, j = 1, \cdots , 4)$ where $\overline{X}_i = \frac{1}{N} \sum_{\alpha=1}^{N} X_{i\alpha}$ is the sample mean of X_i .

From Table 1, we find:

$\overline{X}_1 = 5.48 \quad \overline{X}_2 = 5.56 \quad \overline{X}_3 = 15.55 \quad \overline{X}_4 = 15.49$

$v_{11} = 12.8844 \quad v_{12} = 12.8869 \quad v_{23} = 32.1750 \quad v_{34} = 84.4450$

$v_{22} = 13.2794 \quad v_{13} = 32.0350 \quad v_{24} = 31.5156$

$v_{33} = 87.4000 \quad v_{14} = 31.2381$

$v_{44} = 82.9894 \qquad N = 16$

(2) Compute $\tilde{v}_{ij} = \tilde{v}_{ji}$ $(i, j = 1, \cdots, 4)$ as follows:

$$\tilde{v}_{11} = \tilde{v}_{22} = \frac{v_{11} + v_{22}}{2}; \qquad \tilde{v}_{12} = v_{12} ;$$

$$\tilde{v}_{33} = \tilde{v}_{44} = \frac{v_{33} + v_{44}}{2}; \qquad \tilde{v}_{34} = v_{34} ;$$

$$\tilde{v}_{13} = \tilde{v}_{14} = \tilde{v}_{23} = \tilde{v}_{24} = \frac{1}{4}(v_{13} + v_{14} + v_{23} + v_{24}).$$

In this example,

$$\tilde{v}_{11} = \tilde{v}_{22} = 13.0819 \qquad \tilde{v}_{12} = 12.8869$$
$$\tilde{v}_{33} = \tilde{v}_{44} = 85.1947 \qquad \tilde{v}_{34} = 84.4450$$
$$\tilde{v}_{13} = \tilde{v}_{14} = \tilde{v}_{23} = \tilde{v}_{24} = 31.7409$$

(3) Compute the determinants $|\, v_{ij} \,|$, $|\, \tilde{v}_{ij} \,|$ and their ratio:

$$L_1(vc) = \frac{|\, v_{ij} \,|}{|\, \tilde{v}_{ij} \,|} = y_0 \text{ , say.}$$

We find

$$L_1(vc) = \frac{49.0204}{54.8802} = .893 = y_0 .$$

(4) Evaluate $F(y; 1^b, 2,2; N)$* (given in (A.15)) for $y = y_0$, where in this case $b = 0$. Accept or reject $H_1(vc)$ according as $F(y_0 ; 1^b, 2,2; N)$ is greater than or not greater than α.
For $y_0 = .893$, we find

$F(y_0 ; 1^b, 2,2; N) = .920 \qquad (b = 0, N = 16)$

*5% and 1% points of the sample criteria are tabled in [7].

If we are working at a significance level $\alpha = .05$, we accept $H_1(vc)$.

For testing $H_1(mvc)$, the first step is the same as step 1 above, and the remaining steps are as follows:

(2′) Compute the "pooled" means $\overline{X}_1' = (1/2)(\overline{X}_1 + \overline{X}_2)$,

$$\overline{X}_2' = (1/2)(\overline{X}_3 + \overline{X}_4);$$

then compute $v_{ij}' = v_{ji}'$ $(i, j = 1, \cdots, 4)$, as follows:

$$v_{11}' = v_{22}' = (1/2)(v_{11} + v_{22}) + (N/2)[(\overline{X}_1 - \overline{X}_1')^2 + (\overline{X}_2 - \overline{X}_1')^2];$$

$$v_{33}' = v_{44}' = (1/2)(v_{33} + v_{44}) + (N/2)[(\overline{X}_3 - \overline{X}_2')^2 + (\overline{X}_4 - \overline{X}_2')^2];$$

$$v_{12}' = \tilde{v}_{12} - (N/2)[(\overline{X}_1 - \overline{X}_1')^2 + (\overline{X}_2 - \overline{X}_1')^2];$$

$$v_{34}' = \tilde{v}_{34} - (N/2)[(\overline{X}_3 - \overline{X}_2')^2 + (\overline{X}_4 - \overline{X}_2')^2];$$

$$v_{13}' = v_{14}' = v_{23}' = v_{24}' = \tilde{v}_{13} = \tilde{v}_{14} = \tilde{v}_{23} = \tilde{v}_{24} .$$

In this example,

$$\overline{X}_1' = 5.52, \qquad \overline{X}_2' = 15.52$$

$$v_{11}' = v_{22}' = 13.1075$$

$$v_{33}' = v_{44}' = 85.2091$$

$$v_{12}' = 12.8613$$

$$v_{34}' = 84.4306$$

$$v_{13}' = v_{14}' = v_{23}' = v_{24}' = 31.7409$$

(3′) Compute the determinants $|\, v_{ij} \,|$, $|\, v_{ij}' \,|$ and their ratio

$$L_1(mvc) = \frac{|\, v_{ij} \,|}{|\, v_{ij}' \,|} = u_0 \text{ , say.}$$

We find

$$L_1(mvc) = \frac{49.0204}{71.9514} = .681 = u_0 .$$

(4′) Evaluate $F(u;\ 1^b,\ 2,2;\ N)$ given in (A.13) for $u = u_0$, where in this case $b = 0$. Accept or reject $H_1(mvc)$ according as $F(u_0\ ;\ 1^b,\ 2,2;\ N)$ is greater than or not greater than α.

For $u_0 = .681$, we find

$$F(u_0\ ;\ 1^b,\ 2,2;\ N) = .653 \qquad (b = 0,\ N = 16).$$

In this case also we are led to accept $H_1(mvc)$.

The procedure for testing $H_1(m)$ is as follows:

(1″) Compute $\tilde{v}_{11}$, $\tilde{v}_{12}$, $\tilde{v}_{33}$, $\tilde{v}_{34}$ (see (2) above) and v'_{11} , v'_{12} , v'_{33} , v'_{34} (see (2′) above).

(2″) Compute

$$L_1(m) = \frac{(\tilde{v}_{11} - \tilde{v}_{12})(\tilde{v}_{33} - \tilde{v}_{34})}{(v'_{11} - v'_{12})(v'_{33} - v'_{34})} = z_0 \text{ , say.}$$

(3″) Evaluate $F(z; 1^b, 2,2; N)$ (given in (A.9)) for $z = z_0$, where in this case $b = 0$. Accept or reject $H_1(m)$ according as $F(z_0 ; 1^b, 2,2; N)$ is greater than or not greater than α.
For $z_0 = .763$, we find

$$F(z_0 ; 1^b, 2,2; N) = .140 \qquad (b = 0, N = 16).$$

Accordingly we accept $H_1(m)$ at the 5% level of significance.

On the basis of the above calculations the sample of measurements on 16 rabbits is consistent with any of the three hypotheses, $H_1(mvc)$, $H_1(vc)$ and $H_1(m)$. In other words, the experimenter has no evidence in his sample to indicate that any of the conditions exists which would make him hesitate to use the right and left leg corresponding muscle weights interchangeably. Thus as far as the statistics go, he has no need to fear that the substitution of the weight of a muscle in the untreated leg for the initial weight of the corresponding muscle in the frozen leg will lead to erroneous results. In a broader sense it might be said that the error involved in making the substitution is on the average no greater than the variation observed from rabbit to rabbit. In short, his procedure seems entirely valid.

APPENDIX

A.1. *General Expressions of the Compound Symmetry Hypotheses*

The hypothesis expressed in (1.1) (and in (2.1)) can be generalized as follows. Suppose there are q physical factors in an experiment, where b of the factors are measured once each and the $(b + a)$th factor is measured n_a times ($a = 1, \cdots, h$; $b + h = q$; $n_a \geq 2$). The number of chance quantities would then be $b + n_1 + \cdots n_h = t$, say, and they would fall into $b + h$ groups. A convenient notation for indicating the grouping is* $(1^b, n_1 , \cdots , n_h)$. [For example, the grouping $(1^2, 2,3)$ would mean that there were seven variates X_1 , X_2 , $\cdots$, X_7 , falling into four groups as follows: (X_1), (X_2), (X_3 , X_4), (X_5 , X_6 , X_7). In (1.1) the grouping is* (2,3)]. 'Stability' can be interpreted as meaning:

*The simpler notation $(n_1 , \cdots , n_h)$ is used when $b = 0$; if, moreover, $n_1 = \cdots = n_h = n$, say, the notation (n^h) is sometimes used. If $b = 1$, 1^b is replaced by 1 in the proposed notation.

(1) for each group all means in the group are equal, all variances in the group are equal, and all correlations between distinct chance quantities in the group are equal; and (2) between any two distinct groups all correlations are equal. This 'stability' hypothesis will be represented by $H_1(mvc)$. The hypothesis obtained by removing from $H_1(mvc)$ the condition on the means (see (1)) will be represented by $H_1(vc)$. A third hypothesis, $H_1(m)$, presupposes that $H_1(vc)$ is true and asserts that the condition on the means is true.

The hypothesis expressed in (1.2) (and in (2.2)) can be generalized similarly; however, the restriction for this case that all sets of times be identical requires that $b = 0$ and that $n_1 = n_2 = \cdots = n_h = n$, say.* The generalization of (1.2) (and (2.2)) would leave (1) above unchanged but would substitute the following statement for (2); between any two distinct groups the correlations between simultaneously measured chance quantities are equal and all remaining correlations are equal. (For (1.2) and (2.2) the grouping is (2^2)). This 'stability' hypothesis will be represented by $\overline{H}_1(mvc)$. Hypotheses $\overline{H}_1(vc)$ and $\overline{H}_1(m)$, which are similar to $H_1(vc)$ and $H_1(m)$, respectively, will also be considered. In the remainder of this section a more general method of describing the six hypotheses will be given.

When the grouping of $(X_1, \cdots, X_t)$ is $(1^b, n_1, \cdots, n_h)$ and $H_1(mvc)$, $H_1(vc)$ or $H_1(m)$ is true, the variance-covariance matrix $||A^{ij}||$ of $(X_1, \cdots, X_t)$ is as given in (A.1) (the dashed lines in (A.1) indicate the grouping $(n_1, n_2, \cdots, n_h)$ of the last $t - b$ rows and the last $t - b$ columns of $||A^{ij}||$ (see (2.1), (2.2), and (2.3)):

The hypothesis $H_1(vc)$ asserts merely that $||A^{ij}||$ is as given in (A.1). $H_1(mvc)$ asserts that $H_1(vc)$ is true and that for every a the true means in the $(b + a)$th group of means are equal $(a = 1, \cdots, h)$; $H_1(m)$ asserts that this condition on the means holds, given that $H_1(vc)$ is true.

When $\overline{H}_1(mvc)$, $\overline{H}_1(vc)$, or $\overline{H}_1(m)$ is true, then $b = 0$, the grouping is (n^h) $(t = nh)$, and $||A^{ij}||$ can be divided into h^2 $n \times n$ arrays within each of which all diagonal elements are equal and all off-diagonal elements are equal (see (2.2) for the case of a grouping (2^2); also see [2, p. 453]). $\overline{H}_1(vc)$ asserts that $||A^{ij}||$ has such a form; $\overline{H}_1(mvc)$ asserts that $H_1(vc)$ is true and that for every a the true means in the a-th group of means are equal $(a = 1, \cdots, h)$; $\overline{H}_1(m)$ asserts that this condition on the means holds given that $\overline{H}_1(vc)$ is true.

A.2. *General Expressions of Sample Criteria.*

Let $0_N(x_{1\alpha}, \cdots, X_{t\alpha})$ $(\alpha = 1, \cdots, N)$ be a sample of size N.

*See preceding footnote.

$$
\| A^{ij} \| = \left\| \begin{array}{c|c|c|c|c}
\begin{matrix} A^{11}A^{12}\cdots A^{1b} \\ A^{21}A^{22} \\ \cdots \\ A^{b1} \cdots A^{bb} \end{matrix} & \begin{matrix} C^{11} \cdots C^{11} \\ \cdots \\ C^{b1} \cdots C^{b1} \end{matrix} & \begin{matrix} C^{12} \cdots C^{12} \\ \cdots \\ C^{b2} \cdots C^{b2} \end{matrix} & \cdots & \begin{matrix} C^{1h} \cdots C^{1h} \\ \cdots \\ C^{bh} \cdots C^{bh} \end{matrix} \\ \hline
\begin{matrix} C^{11} \cdots C^{b1} \\ \cdots \\ C^{11} \cdots C^{b1} \end{matrix} & \begin{matrix} A^1B^1\cdots B^1 \\ B^1A^1 \\ \cdots \\ B^1 \quad A^1 \end{matrix} & \begin{matrix} D^{12} \cdots D^{12} \\ \cdots \\ D^{12} \cdots D^{12} \end{matrix} & \cdots & \begin{matrix} D^{1h} \cdots D^{1h} \\ \cdots \\ D^{1h} \cdots D^{1h} \end{matrix} \\ \hline
\begin{matrix} C^{12} \cdots C^{b2} \\ \cdots \\ C^{12} \cdots C^{b2} \end{matrix} & \begin{matrix} D^{12} \cdots D^{12} \\ D^{12} \\ \cdots \\ D^{12} \cdots D^{12} \end{matrix} & \begin{matrix} A^2B^2\cdots B^2 \\ B^2A^2 \\ \cdots \\ B^2 \quad A^2 \end{matrix} & \cdots & \begin{matrix} \\ \\ \cdots \\ \\ \end{matrix} \\ \hline
\begin{matrix} \\ \\ \cdots \\ \\ \end{matrix} & \begin{matrix} \\ \\ \cdots \\ \\ \end{matrix} & \begin{matrix} \\ \\ \cdots \\ \\ \end{matrix} & & \\ \hline
\begin{matrix} C^{1h} \cdots C^{bh} \\ \cdots \\ C^{1h} \cdots C^{bh} \end{matrix} & \begin{matrix} \\ \\ \cdots \\ \\ \end{matrix} & & & \begin{matrix} A^hB^h\cdots B^h \\ B^hA^h \\ \cdots \\ B^h \cdots A^h \end{matrix}
\end{array} \right\| \tag{A.1}
$$

Let $\overline{X}_i = \sum_\alpha X_{i\alpha}/N$ and $v_{ij} = \sum_\alpha (X_{i\alpha} - \overline{X}_i)\ (X_{j\alpha} - \overline{X}_j)$.

The sample criterion $L_1(vc) = |\, v_{ij} \,|/|\, \tilde{v}_{ij} \,|\cdot \| \tilde{v}_{ij} \|$ has the same appearance as $\| A^{ij} \|$ in (A.1); the elements of $\| \tilde{v}_{ij} \|$ are averages of elements in $\| v_{ij} \|$. Let $s = 1, \cdots, b$ and $i_a, j_a, i'_a, j'_a = b + \bar{n}_{a-1} + 1, \cdots, b + \bar{n}_a (\bar{n}_a = n_1 + \cdots + n_a; \bar{n}_0 = 0; a = 1, \cdots, h)$; the elements of $\| \tilde{v}_{ij} \|$ may then be expressed as follows:

$$
\begin{aligned}
&\tilde{v}_{ss'} = v_{ss'}\ ;\ \tilde{v}_{si_a} = \sum_{ja} v_{sja}/n_a\ ; \\
&\tilde{v}_{i_ai_a} = \sum_{ja} v_{ja i_a}/n_a = \tilde{v}_a\ , \text{ say}; \\
&\tilde{v}_{i_ai_a} = \sum_{i'_a \neq j'_a} v_{i'_a i'_a}/n_a(n_a - 1) = \tilde{w}_a\ , \text{ say}, (i_a \neq j_a); \\
&\tilde{v}_{i_ai_{a'}} = \sum_{ja\, i'_{a'}} v_{ja i'_{a'}}/(n_a n_{a'}), (a \neq a'; a, a' = 1, \cdots, h).
\end{aligned} \tag{A.2}
$$

The sample criterion $L_1(mvc) = |\, v_{ij} \,|/|\, v'_{ij} \,|\cdot \| v'_{ij} \|$, like $\| \tilde{v}_{ij} \|$, has the same appearance as $\| A^{ij} \|$ in (A.1); moreover, $v'_{ss'} = v_{ss'}$ and $v'_{i_ai_{a'}} = \tilde{v}_{i_ai_{a'}}$ (see (A.2)), but

$$v'_{i_a i_a} = \tilde{v}_a + \frac{N}{n_a} \sum_{j_a} (\overline{X}_{j_a} - \overline{X}'_a)^2 = v'_a \text{ , say,}$$

$$v'_{i_a j_a} = \tilde{w}_a - \frac{N}{(n_a)(n_a - 1)} \sum_{j'_a} (\overline{X}_{j'_a} - \overline{X}'_a)^2 = w'_a \text{ , say, } (i_a \neq j_a), \tag{A.3}$$

where $\overline{X}'_a = \sum_{j_a} \overline{X}_{j_a}/n_a$.

The criterion $L_1(m) = L_1(mvc)/L_1(vc) = |\tilde{v}_{ij}|/|v'_{ij}|$. Determinants like $|\tilde{v}_{ij}|$ and $|v'_{ij}|$, having the appearance of $|A^{ij}|$ (see (A.1)), can be simplified (see (3.3) in [2, p. 451]). From this simplification it follows that

$$L_1(m) = \prod_{a=1}^{h} \{(\tilde{v}_a - \tilde{w}_a)(v'_a - w'_a)^{-1}\} \tag{A.4}$$

(see (A.2), (A.3)). Note that $L_1(m)$ is independent* of b.

When an H hypothesis is under test the grouping is (n^h) $(t = nh)$, (thus $b = 0$). The criterion $\overline{L}_1(vc) = |v_{ij}|/|\bar{v}_{ij}|$, where

$$\bar{v}_{i_a i_a} = \left(\frac{1}{n}\right) \sum_{j_a} v_{j_a i_a} \text{ ; } \bar{v}_{i_a j_a} = \sum_{i'_a, j'_a} \frac{v_{i'_a j'_a}}{n(n-1)}, \ (i_a \neq j_a \text{ ; } i'_a \neq j'_a);$$

$$\bar{v}_{i_a k_{a'}} = \left(\frac{1}{n}\right) \sum_{i'_a, k'_{a'}} v_{i'_a k'_{a'}} \text{ , } (k'_{a'} = i'_a + (a' - a)n;$$

$$k_{a'} = i_a + (a' - a)n) \tag{A.5}$$

$$\bar{v}_{i_a h_{a'}} = \left(\frac{1}{n(n-1)}\right) \sum_{j_a h'_{a'}} v_{j_a h'_{a'}} \text{ , } (h'_{a'} \neq j_a + (a' - a)n; \ h_{a'} \neq i_a$$

$$+ (a' - a)n; \ h_{a'} \text{ , } h'_{a'} = (a' - 1)n + 1, \cdots, a'n).$$

The criterion $\overline{L}_1(mvc) = |v_{ij}|/|\bar{v}'_{ij}|$, where

$$\bar{v}'_{i_a i_a} = \bar{v}_{i_a i_a} + \frac{N}{n} \sum_{j_a} (\overline{X}_{j_a} - \overline{X}'_a)^2$$

$$\bar{v}'_{i_a j_a} = \bar{v}_{i_a j_a} - \frac{N}{n(n-1)} \sum_{j'_a} (\overline{X}_{j'_a} - \overline{X}'_a)^2, \ (i_a \neq j_a),$$

$$\bar{v}'_{i_a k_{a'}} = \bar{v}_{i_a k_{a'}} + \frac{N}{n} \sum_{i'_a k'_{a'}} (\overline{X}_{i'_a} - X'_a)(\overline{X}_{k'_{a'}} - \overline{X}'_{a'})$$

$$\bar{v}'_{i_a h_{a'}} = \bar{v}_{i_a h_{a'}} - \frac{N}{n(n-1)} \sum_{i'_a, k'_{a'}} (\overline{X}_{i'_a} - \overline{X}'_a)(\overline{X}_{k'_{a'}} - \overline{X}'_{a'}).$$

*When $H_1(m)$ is true, the distribution of $L_1(m)$ is independent of b (see Tables 2 and 3).

The criterion $\overline{L}_1(m) = L_1(mvc)/\overline{L}_1(vc)$; this ratio can be greatly simplified also (see (3.5) in [2, pp. 453–454]).

When $H_1(mvc)$ is true, $L_1(m)$ and $L_1(vc)$ are independent; an entirely similar remark holds regarding $\overline{H}_1(mvc)$, $\overline{L}_1(m)$ and $\overline{L}_1(vc)$.

A.3. *Means, Variances, and Distributions of Sample Criteria When the Corresponding Null Hypotheses are True.*

The numbers on the unit interval, $0 \leq x \leq 1$, form the set of possible values of each sample criterion. The exact distribution of each criterion (when the corresponding null hypothesis is true) has been identified as a product of independent beta variates (see [2]). A beta variate, say X, has the c.d.f.:

$$P_r(X \leq x) = [B(P, Q)]^{-1} \int_0^x t^{P-1}(1 - t)^{Q-1}\, dt = I_x(P, Q), \qquad \text{(A.6)}$$

where $P, Q > 0$ and

$$B(P, Q) = \int_0^1 t^{P-1}(1 - t)^{Q-1}\, dt = \Gamma(P)\Gamma(Q)/\Gamma(P + Q). \qquad \text{(A.7)}$$

$I_x(P, Q)$ is termed the Incomplete Beta Function ratio (see [3], [4]). The mean and variance of X are, respectively:

$$P/(P + Q),\ PQ/[(P + Q)^2(P + Q + 1)]. \qquad \text{(A.8)}$$

When N is large, the distribution of $-N \log_e$ (criterion) has approximately a χ^2-distribution with a number of degrees of freedom that depends on the grouping, $(1^b, n_1, \cdots, n_h)$, (see [2, p. 467]). Small values of the criterion are significant; thus, large values of $-N \log_e$ (criterion) are significant.

In formulas for the c.d.f.'s of $L_1(mvc)$, $L_1(vc)$, $L_1(m)$, $\overline{L}_1(mvc)$, $\overline{L}_1(vc)$, $\overline{L}_1(m)$, the independent variables used will be $u, y, z, \bar{u}, \bar{y}, \bar{z}$, respectively. For any grouping $(1^b, n_1, \cdots, n_h)$ and sample size, N, let $F(u; 1^b, n_1, \cdots, n_h; N)$ be the c.d.f.* of $L_1(mvc)$, $\cdots$, $F(\bar{z}; n^h; N)$ be the c.d.f. of $\overline{L}_1(m)$. C.D.F.'s of various criteria for various groupings (when the corresponding null hypotheses are true) are given in Table 2, together with the numbers of degrees of freedom of the χ^2 distributions discussed below (A.8). The functions listed under "C.D.F." in Table 2 are Incomplete Beta Function ratios; however, it should be noted in the table that in some cases not the criterion but a simple function of the criterion is a beta variate (e.g., $L_1(mvc; 1^b, 3)$). The mean and variance of each criterion (or of the simple function of the criterion) in Table 2 can be obtained from (A.6) and (A.8).

*i.e., $P_r(L_1(mvc) \leq u) = F(u; 1^b, n_1, \cdots, n_h; N)$.

TABLE 2.

Criterion-Grouping	C.D.F.	Degrees of Freedom
$L_1(mvc; 1^b, 2)$	$I_u\left(\frac{N-b-2}{2}, \frac{b+2}{2}\right)$	$b+2$
$L_1(mvc; 1^b, 3)$	$I_{u'}(N-b-3, b+3), (u' = \sqrt{u})$	$2b+6$
$L_1(vc; 1^b, 2)$	$I_y\left(\frac{N-b-2}{2}, \frac{b+1}{2}\right)$	$b+1$
$L_1(vc; 1^b, 3)$	$I_{y'}(N-b-3, b+2), (y' = \sqrt{y})$	$2b+4$
$L_1(m; 1^b, n)$	$I_{z'}\left[\frac{(N-1)(n-1)}{2}, \frac{(n-1)}{2}\right],$	
	$(z' = z^{1/n-1})$	$n-1$
$\overline{L}_1(vc; 2^2)$	$I_{\bar{y}'}(N-4, 2), (\bar{y}' = \sqrt{\bar{y}})$	4
$\overline{L}_1(mvc; 2^2)$	$I_{\bar{u}'}(N-4, 3), (\bar{u}' = \sqrt{\bar{u}})$	6
$\overline{L}_1(m; n^2)$	$I_{\bar{z}'}[(N-1)(n-1)-1, n-1],$	
	$[\bar{z}' = (\sqrt{\bar{z}})^{1/(n-1)}]$	$2(n-1)$
$\overline{L}_1(m; 2^h)$	$I_{\bar{z}}\left(\frac{N-h}{2}, \frac{h}{2}\right)$	h

The exact distribution of certain other criteria (when the corresponding null hypotheses are true) are given below:

$$F(z; 1^b, 2, 2; N)$$

$$= \frac{2\pi}{N-1}\left[B\left(\frac{N-1}{2}, \frac{1}{2}\right)\right]^{-2}\left\{z^{\frac{N-1}{2}} + \frac{I_z\left(\frac{N-1}{2}, 2\right)}{2(N+1)}\right.$$

$$+\left(\frac{9}{8}\right)[(N+3)(N+1)]^{-1}I_z\left(\frac{N-1}{2},3\right)$$

$$+\left(\frac{225}{48}\right)[(N+5)(N+3)(N+1)]^{-1}I_z\left(\frac{N-1}{2},4\right)+\epsilon_N,\Bigg\}, \tag{A.9}$$

$$\left(\frac{96}{\pi}\right)[(N+5)(N+3)(N+1)(N-1)]^{-1}\geq\epsilon_N\geq 0.$$

$$F(z;1^b,3,3;N)=(z')^{N-1}$$

$$-(N-1)(z')^{N-1}\log_e(z'),\ (z'=\sqrt{z}). \tag{A.10}$$

$$F(\bar{u};2^3;N)=\bar{u}^{\frac{N-4}{2}}$$

$$+\frac{(N-2)(N-3)(N-4)^2(N-5)(N-6)}{48}\,\bar{u}^{\frac{N-6}{2}}$$

$$\times\left\{\frac{1}{N-6}-\frac{8\bar{u}^{1/2}}{(N-5)}\right. \tag{A.11}$$

$$-\bar{u}\left(\frac{1}{N-6}-\frac{8}{N-5}+\frac{8}{N-3}-\frac{1}{N-2}\right)$$

$$\left.+\frac{8\bar{u}^{3/2}}{(N-3)}-\frac{\bar{u}^2}{N-2}-\frac{6\bar{u}}{N-4}\log_e\bar{u}\right\}.$$

$$F(\bar{y};2^3;N)=I_v\left(\frac{N-6}{2},\frac{5}{2}\right)+3\left[B\left(\frac{N-6}{2},\frac{5}{2}\right)\right]^{-1}(\bar{y})^{\frac{N-6}{2}}$$

$$\times\left\{(1-\bar{y})^{1/2}+\left(\frac{N-6}{3}\right)(1-\bar{y})^{3/2}\right. \tag{A.12}$$

$$+(N-5)(\bar{y})\ \text{arc tanh}\ \sqrt{1-\bar{y}}$$

$$\left.-(N-4)(\bar{y})^{1/2}\ \text{arc cos}\ \sqrt{\bar{y}}\right\}.$$

$$F(u;1^b,2,2;N)=I_u\left(\frac{N-1}{2},\frac{1}{2}\right)$$

$$+\left[B(N-b-4,\,b+3)B\left(\frac{N-1}{2},\frac{1}{2}\right)\right]^{-1}$$
$$\times\left\{\sum_{q=0}^{b+2}(-1)^q C_q^{b+2}B\left(\frac{b+3-q}{2},\frac{1}{2}\right)\right. \tag{A.13}$$
$$\left.\times\frac{u^{\frac{N-b-4+q}{2}}}{(N-b-4+q)}I_{1-u}\left(\frac{1}{2},\frac{b+3-q}{2}\right)\right\},$$

where $C_q^{b+2}=(b+2)!/[(q!)(b+2-q)!]$.

$$F(u;2,3;N)=I_u\left(\frac{N-3}{2},\frac{3}{2}\right)$$
$$+u^{\frac{N-5}{2}}\left[B\left(\frac{N-3}{2},\frac{3}{2}\right)B(N-5,5)\right]^{-1}$$
$$\times\left\{\left[\frac{8u^{3/2}}{(N-2)}-\frac{4u^{1/2}}{(N-4)}\right]\right. \tag{A.14}$$
$$\times\arccos\sqrt{u}+u\left[\frac{12}{(N-3)}-\frac{u}{(N-1)}\right]$$
$$\times\operatorname{arc\,tanh}\sqrt{1-u}+\left(\frac{2}{3}\right)(1-u)^{3/2}\left(\frac{1}{(N-5)}\right)+u(1-u)^{1/2}$$
$$\left.\times\left[\frac{4}{N-4}-\frac{12}{N-3}-\frac{8}{N-2}+\frac{1}{N-1}\right]\right\}.$$

$$F(y;1^b,2,2;N)=I_v\left(\frac{N-2}{2},\frac{1}{2}\right)$$
$$+\left[B(N-b-4,\,b+2)B\left(\frac{N-2}{2},\frac{1}{2}\right)\right]^{-1}$$
$$\times\left\{\sum_{q=0}^{b+1}(-1)^q C_q^{b+1}B\left(\frac{b+2-q}{2},\frac{1}{2}\right)\right. \tag{A.15}$$
$$\left.\times\frac{y^{\frac{N-b-4+q}{2}}}{(N-b-4+q)}I_{1-v}\left(\frac{1}{2},\frac{b+2-q}{2}\right)\right\}.$$

$$F(y; 2, 3; N) = y^{\frac{N-3}{2}}$$

$$+ \frac{(N-2)(N-3)^2(N-4)(N-5)}{12} y^{\frac{N-5}{2}} \left\{ \frac{1}{N-5} - \frac{6y^{1/2}}{N-4} \right. \quad \text{(A.16)}$$

$$\left. - y\left(\frac{1}{N-5} - \frac{6}{N-4} + \frac{2}{N-2} \right) + \frac{2y^{3/2}}{N-2} - \frac{3y \log_e y}{N-3} \right\}.$$

$$F(z; 1^b, 2, 3; N) = I_z\left(\frac{N-1}{2}, \frac{1}{2} \right)$$

$$+ 2\left[B\left(\frac{N-1}{2}, \frac{1}{2} \right) \right]^{-1} z^{\frac{N-1}{2}} \operatorname{arc\,tanh} \sqrt{1-z}. \quad \text{(A.17)}$$

The means and variances of these criteria (when the corresponding null hypotheses are true) for the cases covered in (A.9) through (A.17) are given in Table 3, together with the numbers of degrees of freedom of the χ^2-distributions discussed below (A.8).

REFERENCES

[1] S. S. Wilks, "Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution", *Ann. Math. Stat. 17*, 257–281.

[2] D. F. Votaw, Jr. "Testing compound symmetry in a normal multivariate distribution", *Ann. Math. Stat. 19*, 447–473.

[3] K. Pearson, *Tables of the Incomplete Beta Function*, Cambridge University Press, 1934.

[4] C. M. Thompson, "Tables of percentage points of the Incomplete Beta Function", *Biometrika, 32*, II, 151–181.

[5] S. S. Wilks, *Mathematical Statistics*, Princeton University Press, Princeton Univ., 1943.

[6] D. F. Votaw, Jr., "Compound symmetry tests in the multivariate analysis of medical experiments", *Biometrics, 5*, 81, 1949 *abstract.*

[7] M. B. Danford, J. A. Rafferty, D. F. Votaw, Jr., "Critical values of certain compound symmetry criteria", *U.S.A.F. School of Aviation Medicine, Project 21-02-033*, Report No. 4 (in press).

TABLE 3.

Criterion, Grouping, Sample Size	Mean	Variance	Degrees of Freedom
$L_1(mvc; 1^b, 2, 2; N)$	$\frac{(N-b-3)(N-b-4)}{N^2}$	$\frac{(N-b-3)(N-b-4)}{N^4(N+2)^2}[2N^2(2b+7) - 4N(b^2+5b+5) - 4(b+3)(b+4)]$	$2b+7$
$L_1(mvc; 2, 3; N)$	$\frac{(N-3)(N-4)(N-5)}{N^2(N+1)}$	$\frac{2(N-3)^2(N-4)(N-5)(13N^3-3N^2-106N-120)}{N^4(N+1)^2(N+2)^2(N+3)}$	13
$L_1(vc; 1^b, 2, 2; N)$	$\frac{(N-b-3)(N-b-4)}{(N-1)^2}$	$\frac{2(N-b-3)(N-b-4)}{(N-1)^4(N+1)^2}\cdot[N^2(2b+5) - 2N(5b+6) - (2b+5)]$	$2b+5$
$L_1(vc; 2, 3; N)$	$\frac{(N-3)(N-4)(N-5)}{N(N-1)^2}$	$\frac{4(N-3)^2(N-4)(N-5)(5N^3-11N^2-20N-10)}{(N-1)^4N^2(N+1)^2(N+2)}$	10
$L_1(m; 1^b, 2, 3; N)$	$\frac{(N-1)^2}{N(N+1)}$	$\frac{2(N-1)^2(3N^2+4N-3)}{N^2(N+1)^2(N+2)(N+3)}$	3

TABLE 3—*Continued*

Criterion, Grouping, Sample Size	Mean	Variance	Degrees of Freedom
$L_1(m; 1^b, 2, 2; N)$	$\frac{(N-1)^2}{N^2}$	$\frac{4(N-1)^2(N^2+N-1)}{N^4(N+2)^2}$	2
$L_1(m; 1^b, 3, 3; N)$	$\frac{(N-1)^2}{N^2}$ *	$\frac{(N-1)^2(2N^2-1)}{N^4(N+1)^2}$ *	4
$\overline{L}_1(mvc; 2^3; N)$	$\frac{(N-4)(N-5)(N-6)}{N(N-1)(N-2)}$	$\frac{24(N-4)^2(N-5)(N-6)(N^2-4N-2)}{(N-2)^2(N-1)^2N^2(N+1)(N+2)}$	12
$\overline{L}_1(vc; 2^3; N)$	$\frac{(N-4)(N-5)(N-6)}{(N-1)(N-2)(N-3)}$	$\frac{18(N-4)^2(N-5)(N-6)(N^2-5N+2)}{N(N-1)^2(N-2)^2(N-3)^2(N+1)}$	9

*The mean and variance here are for the square root of the criterion.

A SIMPLE TREND TEST WITH APPLICATION TO ERYTHROCYTE SIZE DATA

G. Elfving and J. H. Whitlock[1]

Cornell University[2]

1. *Summary.* A simple rank statistic is used to detect a trend in erythrocyte size data; the method may probably prove useful in other cases where a more refined regression analysis would be too laborious. The proposed criterion is essentially Kendall's S (or τ) pooled over several sets of observations. From the statistic in question, an estimate of slope may be derived. The efficiency of this estimate may serve as a measure of the efficiency of the test. It is calculated for the case that the observations refer to equidistant time points and are normally distributed around regression lines of constant slope.

A short description of the practical procedure is found at the end of section three.

2. Although sheep erythrocytes are only one third as large (27 cu. microns) as those of man (90 cu. microns), the data which we have accumulated over the past four years indicates that the normal variance in volume is of the same order of magnitude. Diagnosis of several dietary deficiencies rests upon determinations of changes in both shape and size of erythrocytes. It is obvious that a change of 10 cubic microns in average cell volume represents a much more severe disorder in sheep than in man. Accordingly, it becomes desirable to diagnose such changes before they become lethal. The blood samples from anemic sheep used in our experiments (Whitlock, 1949) were tested by turbidimeter (Whitlock, 1947) with a red and green filter as well as being subjected to routine erythrocyte count and hematocrit determinations. The erythrocyte count and turbidimetric determinations were made on the same diluted sample of blood. Another sample was used for the hematocrit determinations. The red filter turbidimetric reading could be translated into expected counts and expected hematocrits.

[1]The method was first suggested to the latter author by Professor W. Feller of the Mathematics Department, Cornell University. The authors are also indebted to Professor J. W. Tukey, Princeton University, for improvements in the computing scheme.—Research supported in part by ONR.

[2]Mathematics Department and New York State Veterinary College, Cornell University.

The difference between actual and expected counts and hematocrits gave us a measure of cell volume and cell shape. The difference between the red and green filter turbidimetric readings was translated into hemoglobin values. Cell volume also could be determined by the ratio between the hematocrit and the cell count. With two sub-samples, our technic gave us two determinations of cell volume, one of cell shape, and one of hemoglobin concentration per cell. The determination of red cell count, hematocrit, and hemoglobin as routinely done in most laboratories requires three sub-samples and yields one determination of cell volume and one of hemoglobin concentration per cell. In other words, the turbidometric technic yields more information from fewer sub-samples.

Hematocrit determinations are usually reported as packed cell volume percentage and are probably only accurate within about $\pm 5\%$. The erythrocyte count error lies between 9 and 16%. When errors are so large in comparison to the effects expected, all animals in an experimental group would not show a uniform trend in measured cell volume even if a well defined trend were present. The essential problem, therefore, is to determine whether or not a trend in a series of measurements of a group of animals is sufficiently pronounced that mere chance would not account for its appearance. Regression technics would solve the problem, but they are too laborious to be worthwhile.

3. Under these circumstances, a rank method was adopted. Beside being quick, it has the advantage that the probability distribution (and thus, particularly, the percentage points) of the test criterion are independent of any assumptions about the distribution of the observed values as long as no trend is present.

Table I lists data from previously reported experiments (Whitlock, 1949) indicating determinations of erythrocyte sizes in various anemic sheep. To the size values x_i $(i = 1, \cdots, n)$ of each individual i, rank numbers (in parentheses) are assigned in order of increasing magnitude;[1] e.g., for sheep number 32, the cell volumes, in order are:

27.5	29.0	31.3	28.0

and the corresponding rank numbers:

(R)	1	3	4	2

If there is a strong positive (or negative) trend in the x-value, the rank numbers may be expected to occur nearly in the natural order $1, 2, \cdots, n$ (or $n, \cdots 2, 1$, respectively). In order to measure the degree of dis-

[1] If two or more x-values coincide, it is usual to assign to all of them the average of the corresponding rank numbers.

TABLE I

Sheep Number	Determination of Cell Size (Cubic Microns)					
	1st	2nd	3rd	4th	5th	k
25	25.4 (1)	26.5 (2)				0
63	24.1 (1)	24.9 (2)				0
26	26.3 (2)	24.1 (1)				2
55	27.3 (1)	29.0 (2)	29.4 (3)			0
39	23.7 (1)	28.2 (2)	28.4 (3)			0
66	25.0 (1)	25.7 (2)	29.6 (3)			0
47	27.5 (1)	27.9 (2.5)	27.9 (2.5)			1
6	23.0 (1)	28.4 (3)	25.0 (2)			2
75	25.2 (1)	26.6 (2)	27.8 (3)			0
5	24.1 (2)	22.9 (1)	28.9 (3)	30.1 (4)		2
35	21.3 (1)	23.6 (2)	25.8 (3)	29.4 (4)		0
32	27.5 (1)	29.0 (3)	31.3 (4)	28.0 (2)		4
45	25.4 (1)	32.5 (3)	33.2 (4)	31.6 (2)		4
13	28.5 (4)	28.1 (3)	26.9 (1)	27.0 (2)		10
74	28.4 (2)	27.2 (1)	29.6 (3)	36.0 (5)	31.1 (4)	4
40	29.6 (1)	30.0 (2)	33.5 (5)	32.0 (3.5)	32.0 (3.5)	5
34	24.9 (1)	26.3 (2)	31.3 (3)	38.0 (5)	32.8 (4)	2

$K = 36$

order in the rank sequence, we count the number of *inverted pairs*, i.e. the number of pairs of rank numbers with the smaller rank to the right of the greater. In our example there are two such pairs, namely (3,2) and (4,2); the remaining pairs (1,3), (1,4), (1,2), and (3,4) are not inverted. Alternatively, the number of inverted pairs is obtained by noting, for each item in (R), the number of smaller items to the right, and adding these numbers. We take *twice* the number of inverted pairs and call it k, the redoubling being to facilitate certain calculations in the following.

The statistic k is equivalent with Kendall's S or τ (Kendall, 1945, I, 391)[1]; in fact, it is easily verified that

$$k = \tfrac{1}{2}n(n-1) - S = \tfrac{1}{2}n(n-1)(1-\tau).$$

If there is no trend, k is a random variable (cf. loc. cit. p. 403) with expectation

[1]Tukey has proposed to call the lesser of k and S the Kendall sum.

$$a_n = \tfrac{1}{2}n(n-1) \tag{1}$$

and variance

$$\sigma_n^2 = \frac{1}{18}n(n-1)(2n+5). \tag{2}$$

The distribution of S is tabulated by Kendall for $n = 2, 3, \cdots, 10$ and we are thus, as far as only a single individual is concerned, in a position to judge of the significance of an observed number of inversions.

In general, however, we want to pool the information obtained from several, say r, individuals. For this purpose, the simplest procedure is to add the single inversion numbers to form a pooled inversion number

$$K = k_1 + \cdots + k_r \tag{3}$$

The exact distribution of K—even under the null hypothesis that no trend is present—is complicated; as soon as r is not very small, the distribution is, however, fairly near to the normal, with mean

$$A = \frac{1}{2}\sum_{i=1}^{r} n_i(n_i - 1) \tag{4}$$

and variance

$$S^2 = \frac{1}{18}\sum_{i=1}^{r} n_i(n_i - 1)(2n_i + 5), \tag{5}$$

where n_i denotes the number of observations on the individual i. We thus simply have to compute A and S and then judge of the significance of K according to standard rules.

TABLE II

n	a_n	$3\sigma^2_n$
2	1	3
3	3	11
4	6	26
5	10	50
6	15	85
7	21	133
8	28	196
9	36	276
10	45	375

The procedure may be facilitated by using Table II which gives a_n and $3\sigma_n^2$ for $n = 2, \cdots, 10$; both are integers. The applier simply has to (a) count the redoubled number k of inverted pairs of rank numbers for each individual, (b) take a_n and $3\sigma_n^2$ from Table II for all occurring numbers n of observation, (c) add these three items over all individuals, thus obtaining the quantities K, A, and $3S^2$, (d) take the square root of $3S^2/3$ thus obtaining the standard deviation, S, of K; (e) divide the difference $K - A$ by S. If the difference $K - A$ is more than twice the standard deviation, the trend may be judged significant. If the difference is more than thrice the standard deviation, the results are highly significant.

The calculations are given in Table III. The resulting normal deviate amounting to more than four times its standard deviation, the trend is clearly significant. It should be noted that we are testing for trend in either direction, thus, the significance level applied to the normal deviate should be that of a two-sided test.

TABLE III

CALCULATION OF NORMAL DEVIATE IN EXAMPLE

Observations per series	Number of series	Number of Series $\times a_n$	Number of Series $\times 3\sigma^2_n$
2	3	3	9
3	6	18	66
4	5	30	130
5	3	30	150
Sums	17	$A = 81$	$3S^2 = 355$
		$K = 36$	3
		$K - A = -45$	$S^2 = 118.3$

$$\text{Deviate} = \frac{-45}{\sqrt{118.3}} = -4.13$$

4. It may be of a certain interest to study the *efficiency* of the test used. For this purpose, we first have to choose an appropriate probability model for the phenomenon investigated, valid for the general case that a trend is present. Of the parameters necessarily involved in this model, one, say θ, has to be a measure of the trend; without loss of generality, we may assume that the value $\theta = 0$ corresponds to the null hypothesis that no trend is present. The statistic k (or the pooled statistic K) yields, suitably transformed, an unbiased estimate of θ. We will calculate the variance of this estimate and compare it to the

theoretic lower bound of the variance of any unbiased estimate of θ obtainable from the original set of observations (cf. e.g. Cramer, 1946, p. 490–497). The ratio of the latter variance to the former measures the efficiency of the estimate of θ and thus, in a certain sense, the efficiency of the test proposed.

The most easily managed probability model is given by the hypothesis that the observed values x are normally distributed, with the mean 0 and standard deviation σ, around straight lines of common slope β; the level of these lines may vary from individual to individual. This hypothesis may be expressed by the formula:

$$x_i(t) = \alpha_i + \beta t + \xi_{it}\,, \tag{6}$$

where the ξ_{it}'s (the pure random effects) are normal $(0, \sigma)$. As a measure of the essential trend, i.e. the trend compared with the standard deviation of the random fluctuations, we introduce $\theta = \beta/\sigma$.

Let us now, to begin with, study the efficiency of a single statistic k for testing the null hypothesis $\theta = 0$. The probability distribution of k is, without difficulty, seen to depend only on θ and the number n of time points, the latter being supposed equidistant. Its analytical form is complicated; we may, however, easily construct an unbiased estimate of θ, based on k and valid for *small* θ's by expanding the expectation of k in a Taylor series

$$E(k) = a_n + b_n\theta + \cdots \tag{7}$$

and neglecting terms of second and higher order. Obviously,

$$t = \frac{k - a_n}{b_n} \tag{8}$$

is an estimate of the kind mentioned. Here, a_n is given by (2); for b_n one obtains, by some calculations here omitted,

$$b_n = -\frac{1}{6\sqrt{\pi}}\, n(n^2 - 1). \tag{9}$$

Thus, according to (3),

$$\text{var}\,(t) = \frac{\sigma_n^2}{b_n^2} = 4\pi\, \frac{n + 5/2}{n(n-1)(n+1)^2}. \tag{10}$$

On the other hand, the lower bound of any unbiased estimate of θ turns out to be

$$\inf \text{var}\,(t) = \frac{12}{n(n^2 - 1)} + \frac{\theta^2}{2n}; \tag{11}$$

hence, for vanishing θ, the efficiency of k is

$$e_n = \frac{3}{\pi} \cdot \frac{n+1}{n+5/2}. \tag{12}$$

We now turn to the efficiency of the pooled inversion number K. Adding (7) over all individuals we get

$$E(K) = A + B\theta + \cdots \tag{13}$$

where $A = \Sigma\, a_{n_i}$, $B = \Sigma\, b_{n_i}$. This yields for small θ's the unbiased estimate $T = (K - A)/B$, with the variance

$$\text{var}\,(T) = \frac{\text{var}\,(K)}{B^2} = \frac{\Sigma\sigma^2_{n_i}}{(\Sigma b_{n_i})^2} = 4\pi\,\frac{\Sigma n_i(n_i-1)(n_i+5/2)}{(\Sigma n_i(n_i^2-1))^2}. \tag{14}$$

The theoretic lower bound of the variance is found to be

$$\inf \text{var}\,(T) = \frac{12}{\Sigma n_i(n_i^2-1)} + \frac{\theta^2}{2\Sigma n_i};^{1} \tag{15}$$

hence, for small θ, the efficiency of K is

$$e = \frac{3}{\pi} \cdot \frac{\Sigma n_i(n_i^2-1)}{\Sigma n_i(n_i-1)(n_i+5/2)}. \tag{16}$$

If all individuals show the same number n of observations, (16) is identical with (12). In this case, the efficiency increases from $e_2 = 2/\pi = 0.637$ to $e_\infty = 3/\pi = 0.955$ as r increases.

If the n_i's are unequal, (16) depends on the relative frequency of the different n_i's. If e.g. the values $n = 2, 3, 4, 5$ occur with equal frequency, e is found to be 0.743. The efficiency could in the case of varying n_i's be somewhat improved by using a weighted sum instead of the straight sum $K = \Sigma\, k_i$. The improvement seems, however, to be too small to justify the more complex computations. In the numerical example just mentioned, the improved efficiency is found to be 0.744.

REFERENCES

Cramér, H. *Mathematical methods of statistics*. Princeton University Press, 1946.

Kendall, M. G. *The advanced theory of statistics, 2 ed.* Ch. Griffin and Co., London, 1945.

Whitlock, J. H. The use of photo-electric turbidometry in the determination of red cell counts, hematocrits, and hemoglobin. *Blood*, II, 463–473, 1947.

Whitlock, J. H. The relationship of nutrition to the development of the trichostrongylidoses. *Cornell Vet.* XXXIX, 146–182, 1949.

Wilks, S. S. *Mathematical statistics*. Princeton University Press, 1943.

[1] As soon as the total number of observations is reasonably large, this limit is closely attained by the estimate $_\theta = {}_\beta/_\sigma$, where $_\beta$ and $_\sigma$ are the standard estimates of β and σ provided by the theory of linear regression; cf. Wilks (1943), Ch. 8.

A NOTE ON THE FOUR BY FOUR LATIN SQUARES

W. J. Youden

National Bureau of Standards, Washington, D. C.

It has been pointed out by R. A. Fisher (1) that randomized blocks have the advantage that it is possible to isolate the appropriate components of error applicable to any specified comparison of the treatments. This is useful when there is any reason to question the use of the pooled error term.

The Latin Square effects a double elimination of block differences and there appears to be no reference to separating the residual interaction of rows and columns into components which could be associated with the comparisons among the letters which designate the treatments in a Latin Square. It is possible to achieve this segregation for the Second Transformation Set of 4 x 4 Latin Squares (2).

The standard square of this set is

A_1	B_2	C_3	D_4
B_3	A_4	D_1	C_2
C_4	D_3	A_2	B_1
D_2	C_1	B_4	A_3

The orthogonal subscripts are used to distinguish the various replications. Inspection shows that if, e.g., the comparison of A and C is of particular interest the error term belonging to this comparison is obtained by taking the following two comparisons of A and C, since these contrasts are independent of row and column effects.

$$(A_1 + A_2) - (C_3 + C_4) = d_1$$

and

$$(A_4 + A_3) - (C_2 + C_1) = d_2$$

The discrepancy between these two estimates of twice the difference between A and C constitutes one of the six degrees of freedom available for the error term in a 4 x 4 square. Simply square the quantity $(d_1 - d_2)$ and divide by 8. The other five degrees of freedom corresponding to the remaining five contrasts among the four letters A, B, C and D may be individually obtained in the same way. A simple numerical example will reveal that the sum of the squares for the six individual degrees of freedom isolated in this way do total to the sum of squares for error obtained in the usual way.

Obviously one degree of freedom for error isn't very helpful. It will sometimes happen, as in a recent experiment on the thickness of protective coatings on metals that a number of replications (in this case eight) of the Latin Square are available and the individual degrees of freedom from each square may be accumulated. There was, in this instance, a real reason for isolating these individual components in the error. The "treatments" were measurements made by four different laboratories and prior experience had shown that the laboratories did not achieve the same precision in the measurements. This arrangement made it possible to determine and appraise the bias between laboratories.

This feature of the 4 x 4 squares may also be a helpful instruction aid since it makes possible a direct and easily understood computation of the error term for this Latin Square. It is certainly not apparent to many who use Latin Square designs that the residual sum of squares is the error term.

REFERENCES

(1) Fisher, R. A. *The Design of Experiments.* Oliver and Boyd, Edinburgh, 5th edition, Section 24, 1949.

(2) Fisher, Ronald A. and Yates, Frank. Table XV. *Statistical Tables.* Hafner Publishing Company, Inc., New York, 1948.

AN INVERSE SAMPLING PROCEDURE FOR BACTERIAL PLATE COUNTS

MARTIN SANDELIUS

University of Uppsala and University of Washington

A COMMON PROCEDURE for estimation of bacterial density consists in counting the total number of colonies of bacteria on a circular plate. Since each colony has grown from a single bacterium living on the medium used on the plate, the result of counting is an estimate of the density of living bacteria per cc of the fluid from which the sample is taken. Experience has shown in many cases that the number of colonies found approximately follows a Poisson distribution.

Often the number of colonies found will be so large that this counting procedure becomes laborious. An alternative procedure would be to apply the principle of inverse sampling, recently adopted in the binomial case (cf. (1) and (2)). We choose on the circular plate a radius vector at random and determine the smallest sector beginning at this vector which contains exactly k colonies, each colony being reduced to a point by considering only its centre, and any colony with its centre on the initial radius vector being excluded. If the plate contains less than k colonies it may be assumed that the sampling process continues on other similar plates, in which case the final sector will have an angle greater than 360°. Further it will be supposed that the quality of the plates is so good that the Poisson distribution of the numbers of colonies within equal sectors is the same.

Choosing the length of the whole circumference of a plate as the unit, we let x denote the length of the sector determined by the inverse sampling procedure. Denoting by a the density of bacteria per cc of the fluid sampled and by v the amount, in cc, of this fluid on each plate, it follows (cf. e.g. (3), Sect. 3) that $2\,avx$ has a chi-square distri-

bution with 2 k degrees of freedom. Assuming k greater than 2, it follows that $(k - 1)/vx$ is an unbiased estimate of a, the variance of this estimate being $a^2/(k - 2)$ (cf. e.g. (4), Sect. 33.3, Ex. 3). Further the interval

$$\left(\frac{\chi^2_{\alpha/2}}{2vx}, \quad \frac{\chi^2_{1-\alpha/2}}{2vx}\right)$$

is a confidence interval for a corresponding to the confidence coefficient $1 - \alpha$, where $0 < \alpha < 1$ and χ^2_p is given by the relation Prob $(\chi^2 \leq \chi^2_p) = p$, χ^2 being taken with 2 k degrees of freedom.

Suppose now that, instead of having the same concentration of bacteria on each plate, different concentrations are used. Then usually 3 or 4 plates will be sufficient to find at least k colonies. Let e.g. the 1st plate contain v cc of the fluid sampled, the 2nd 10 v cc, the 3rd 100 v cc, etc. The same method is still applicable; the only modification necessary is to change the length of the circumference of the 2nd plate to 10, of the 3rd plate to 100, etc.

The author wishes to express his gratitude to Dr. Douglas Chapman, University of Washington, who read the manuscript and made valuable suggestions.

REFERENCES

(1) J. B. S. Haldane, "On a method of estimating frequencies", *Biometrika*, *33*. 222–225, 1945.
(2) D. J. Finney, "On a method of estimating frequencies", *Biometrika*, *36*, 233–234, 1949.
(3) S. Malmquist, "A statistical problem connected with counting of radioactive particles", *The Annals of Mathematical Statistics*, *18*, 255–264, 1947.
(4) H. Cramér, *Mathematical Methods of Statistics*, Princeton, 1946.

QUERIES

80 QUERY: In connection with partial regression and multiple co-variance I have a question which disturbs me although it probably has a very simple explanation. In carrying out the calculations we find the following:

$$Sx_{1y} = -310.12$$

$$Sx_{2y} = -660.24$$

$$Sx_{3y} = 10{,}462$$

The partial regression coefficients are as follows:

$$byx_1 \cdot x_2x_3 = 0.35313$$

$$byx_2 \cdot x_1x_3 = -0.95405$$

$$byx_3 \cdot x_1x_2 = 0.33149$$

As I understand it, one may state that the product $(byx_1 \cdot x_2x_3)(Sx_{1y})$ may be considered as expressing the portion of the sum of squares of Y that are associated with the variations in X_1. This will, however, obviously be a minus number. The other two are positive values. I would greatly appreciate if you could explain to me the interpretation of the negative values when they are obtained.

ANSWER: It is unfortunate, but multiple regression does not afford a nice method for separating the sum of squares into orthogonal batches attributable to the several variables.

Perhaps you can get the information you wish by looking at the standard partial regression coefficients with their associated standard errors. Or, you might use the following device which yields an exact test. Analyze the variance (a regression with 3 independent variables is used as illustration) in this way:

Source of Variation	Degrees of Freedom	Sum of Squares
2 Independent Variables	$n - 3$	$(1 - R_{Y.12}{}^2)Sy^2$
3 Independent Variables	$n - 4$	$(1 - R_{Y.123}{}^2)Sy^2$
Difference	1	$(R_{Y.123}{}^2 - R_{Y.12}{}^2)Sy^2$

The difference with 1 degree of freedom is orthogonal to the sum of squares immediately above it, so $F = (R_{Y.123}{}^2 - R_{Y.12}{}^2)(n - 4)/$

$(1 - R_{Y.123}^2)$ tests the hypothesis that the two σ's are the same. Rejection of the hypothesis leads to the conclusion that the third independent variable is, in the population, individually associated with Y.

81 **QUERY:** In a preliminary experiment on the percentage vitamin content of replicate samples at different age levels the following results were obtained:

PERCENTAGE VITAMIN CONTENT OF REPLICATE SAMPLES AT DIFFERENT AGE LEVELS

Age (Days)	Vitamin Content (%)	Means
0	36, 29, 30, 32	31.8
3	47, 11, 50, 55	40.8
7	24, 57, 57, 59	49.3
10	21, 59, 55, 50	46.3
14	56, 54, 51, 33	48.5
17	53, 31, 59, 59	50.5
21	73, 72, 42, 64, 62, 64	62.8

REGRESSION ANALYSIS

Source of Variation	d.f.	S.S.	M.S.	F
Between Age Groups	6	2630	438	2.25
Regression	1	2249	2249	11.53**
Deviations	5	381	76	
Within Age Groups	23	4474	195	
Total	29	7104		

I would like you to comment on the following:

(1) Are we justified in doing a regression analysis when the F ratio for "between age groups" is not significant?

(2) We have reason to believe that the data are best represented by a cubic equation, yet we get no significant deviation from linear regression. Can this be explained on the basis of an excessively large error term?

(3) We are planning to repeat this experiment using eight equally spaced age levels (0, 3, 6, 9, 12, 15, 18, 21). The samples are from a natural product and the assays are expensive and time consuming. We are contemplating an increase to 10 or 15 replicates at each age level. Would you consider this a sufficiently large sample in view of the large intrinsic error?

ANSWER: (1). Yes. The form of the analysis is inherent in your experiment and could have been incorporated in your project outline.

(2) There are four ways in which the non-significant deviation from linearity might be accounted for. (i) The population regression may be straight—this is the evidence furnished by the sample. (ii) The curvilinearity in the population may be too small to be detected by a sample of any feasible size. (iii) The sample drawn may be too small to detect an existing curved regression. (iv) Your sample may be one of the unusual kind that happens to show small deviations though drawn from a population with detectible curvilinearity.

(3) This is an appropriate question but it cannot be answered from the evidence given. If you will tell me, either graphically or algebraically, the details of the cubic you expect, I might be able to estimate the size of sample required by use of your sample deviations from the expected regression.

P.S.: The reply to this said, "This would be clearer to you if I could reveal the true nature of the data, which unfortunately I cannot." So, that's that.

82 QUERY: The question has arisen as to whether the analysis of variance procedure may be applied to data which are not replicated. The data consist of 225 items as follows:

The material consisted of 3 sweet corn varieties. Each variety was harvested at 3 predetermined stages of maturity as indicated by an objective test. The field replicates were combined, and each variety at each stage of maturity was divided into 5 lots, each being a different temperature treatment. The lots were then analyzed for certain chemical components at 5 successive periods of time. Since the zero period of time could not be subdivided into different temperatures, we are planning to repeat the value for this first period five times in the calculations. This should reduce the degrees of freedom from 224 to 188.

Is there any reason why each set of chemical determinations cannot be analyzed by the variance methods?

ANSWER: The superficial answer to your question is "No". There is no reason why the arithmetic of analysis of variance may not be done in any one of many ways. The pertinent question is this: "Under what circumstances will the analysis of variance produce meaningful results?" The conditions for such results were clearly stated by Eisenhart in this journal, Vol. 3, pages 1–21, 1947.

I shall assume that the collections of material from the original ran-

domized blocks experiment were taken at random positions within the plots, equal amounts from each plot, so that the environmental effects were properly incorporated in the harvested corn: it may be that all the corn from each plot was included. I assume also that the harvested material was randomly assigned to the lots for temperature treatments. This would produce 45 lots which would presumably meet all specifications for a meaningful analysis.

The 5 (or 4) chemical determinations on a single lot can scarcely be expected to be independent, but deviations from linear regression might be. If not, the second degree regression might have some meaning and deviations from it would probably be random. If your time intervals are equal and if you had run your analysis on each lot at the initial time, it would be easy to analyze the variance, perhaps as follows:

Variety, V,	2 degrees of freedom
Stage of Maturity, M,	2
VM	4
Temperature, T,	4
TV	8
TM	8
TVM	16
Linear Trend, L,	1
LV	2
LM	2
LT	4
LVM	4
LTV	8
LTM	8
$LTVM$	16
Remainder	135
Total	224

The analysis with 5 determinations on some lots and 4 on others would be messy. I think your plan of repeated use of the initial determination in each lot will not introduce much inaccuracy into the regressions. I say this partly because, with only three replications of the varieties, great precision cannot be expected.

If theoretical considerations and examination of your data suggest that curved regressions are required, and if this is verified by a large mean square in "Remainder", you can extend the analysis of variance to include second degree terms.

83 QUERY: Using wooden blocks secured from the sapwood of a single log, a decay experiment was conducted in eight decay chambers. Each chamber originally contained eight blocks, but

one of these was removed from each chamber at the end of each 2 week period until all the blocks had been removed. This gave an 8 by 8 experiment with 8 different decay chambers and 8 different decay periods. The variable being studied was weight loss due to decay expressed as a percent of the original oven dry weight. It turned out that both the decay chambers and the length of the decay periods had a significant effect on the percent weight loss due to decay. In addition it was desired to see if the original specific gravity of each piece had any effect on the resistance of the piece to decay. In answering this latter question I resorted to an analysis of covariance. For the wood decay organism *Polyporus versicolor*, I got the following covariance of original specific gravity times $10^2 (= x)$ and the percent weight-loss due to decay $(= y)$:

Source of variation	d.f.	Sx^2	Sxy	Sy^2	Errors of Estimate		
					Sum of sq.	d.f.	Mean Square
Decay chambers	7	6.86	−11.48	1113.36			
Decay periods	7	7.11	60.64	2887.36			
Error	49	134.26	20.98	909.76	906.39	48	18.88
Total	63	148.23	70.14	4910.48			

Error regression coefficient, $b = 0.1563$

Student's "t" $= 0.1563 \sqrt{\frac{134.26}{18.88}} = 0.417$

Since this value of "t" does not even approach significance, I assume that this experiment does not indicate any relationship between original specific gravity and decay resistance of the wood when freed of the effect of decay chambers and decay periods.

Is my interpretation of the meaning of error regression correct?

Is it correct to calculate, test, and interpret the error regression coefficient as I have done?

ANSWER: Your interpretation of the covariance is correct. All the evidence points to the conclusion that the regressions in decay chambers and decay periods, as well as in error, may be no more than sampling variation in samples drawn from a population whose regressions are all zero.

84 QUERY: We frequently must compare percentages, e.g. of mortalities, of a treated and an untreated group of animals in toxicological research. Usually the total number of animals in each group is less than 30. We have used 4 methods for this comparison. These are:

(1) Calculation of critical ratio (*C.R.*) as in Garrett's *Statistics in Psychology and Education*, Longmans, Green and Co. (1947) p. 218, where

$$C.R. = \frac{\%_{01} - \%_{02}}{100\sqrt{\frac{p_1 q_1}{N_1} + \frac{p_2 q_2}{N_2}}}$$

The probability associated with the *C.R.* was obtained from a table of fractional parts of the total area under the normal probability curve.

(2) Calculation of the uncorrected chi-square.

(3) Calculation of the corrected chi-square as in Snedecor's *Statistical Methods*—Iowa State College Press (1946).

(4) Calculation of *t*-test using "Student's" distribution where

$$t = \left(\frac{p_1 - p_2}{\sqrt{\Sigma d_1^2 + \Sigma d_2^2}}\right)\sqrt{\frac{N_1 N_2}{N_1 + N_2}(N_1 + N_2 - 2)}$$

and $\sum d^2 = \sum (x - \bar{x})^2$ with death = 1, survival = 0.

A list of the probabilities associated with these 4 methods on several sample calculations is given below.

Mortalities	Probability Value Associated with			
	C.R.	Uncorr. Chi square	Corrected Chi square	*t*-test
1/30 vs. 5/30	0.08	0.04	0.20	0.23
5/30 vs. 9/30	0.22	0.22	0.37	0.42
10/30 vs. 15/30	0.18	0.20	0.30	0.37
1/20 vs. 5/20	0.06	0.08	0.18	0.08
5/20 vs. 9/20	0.18	0.18	0.31	0.22
1/10 vs. 2/10	0.53	0.52	——	0.55
2/10 vs. 5/10	0.14	0.17	0.35	0.17

We realize that a small difference between probability values is of little significance. The first two methods, *C.R.* and uncorrected chi square, seem to yield comparable results; with the results of the *t*-test sometimes similar and often quite different. Furthermore, the corrected chi-square test always yields results closer to a p of 1.0 than the first

two methods. Our question is: is one of the 4 methods preferable to the others for this type of data when N is small?

ANSWER: Yates has showed (*Journal of the Royal Statistical Society Supplement, I*, 217–235, 1934) that, for samples with a single degree of freedom drawn from a binomial population, a simple correction for continuity gives a satisfactory approximation to the continuous value of chi-square. Since your samples are small, the use of corrected chi-square is, on the average, a better approximation to continuous chi-square than the other methods described.

As for the normal approximation (your first method), Student found *Biometrika, VI*, 1–25, 1908) that this consistently underestimates the probability in the regions ordinarily used for rejection. Your first column would make this evident were it not for mistakes in computation, coupled with the fact that your probabilities are large.

The quantity which you have called t does not follow the Student—Fisher distribution.

Since the tabulated t with infinite degrees of freedom (the normal distribution) is the same as that of chi with one degree of freedom, the identical probability can be read from the table of chi-square and from the (two-tailed) table of the normal distribution if t is calculated as follows:

$$t = \frac{p_1 - p_2}{\sqrt{2\bar{p}\bar{q}/n}}$$

Here, $\bar{p} = (p_1 + p_2)/2$ and n is the size of each sample.

It happens that testing the null hypothesis is not particularly interesting in any of your examples because P is greater than 0.05 in all of them. I took mortalities 1/20 vs. 6/20 and calculated these probabilities:

(1) Normal Approximation, $P = 0.028$
(2) Chi-square (or t correctly calculated), $P = 0.038$
(3) Corrected Chi-square, $P = 0.096$

For interest, I also calculated the exact test (R. A. Fisher, *Statistical Methods for Research Workers*, section 21.02) and got $P = 0.046$, which happens in this sample to be nearest the probability turned up by uncorrected chi-square. In some of your samples the exact test is as easily calculated as chi-square. In others, having a small expectation in one cell of the contingency table, a device given by Yates yields close approximations. This device is described in connection with Table VIII in *Statistical Tables* by Fisher and Yates.

THE BIOMETRIC SOCIETY

Regional Officers. The following results of regional annual elections for 1950 should be recorded: Australasian Region—Vice President, Dr. E. A. Cornish; Secretary-Treasurer, John Keats. British Region—Vice President, Prof. R. A. Fisher; Secretary, Dr. D. J. Finney; Treasurer, Dr. A. R. G. Owen. Eastern North American Region—Vice President, Dr. Joseph Berkson; Secretary-Treasurer, Prof. Walter T. Federer; Committee term 1950–1952, Dr. W. J. Youden, Lila F. Knudson. Indian Region—Vice President, Dr. P. C. Mahalanobis; Secretary, Dr. C. Radhakrishna Rao; Treasurer, Anakul Chandra Das.

Proceedings of a Biometrical Clinic on Entomological Problems. One of the objectives of the Biometric Society is the "dissemination of effective mathematical and statistical techniques". A popular means of implementing this objective takes the form of a jointly sponsored meeting with a biological society. In several cases the program has consisted of questions answered informally by a panel of biometricians, the questions being submitted by the biologists. The proceedings of such a joint session of the American Association of Economic Entomologists and the Eastern North American Region of the Biometric Society held in December 1948 were recorded electronically. After the proceedings were transcribed, edited and reviewed by a number of entomologists and professional biometricians, the Council agreed to their publication as an experiment, on a self-supporting basis. Announcements were included in a general mailing to the members of the two sponsoring organizations and resulted in 98 prepublication orders from the Biometric Society and 489 from the Economic Entomologists.

In consequence 700 copies were ordered and the bound proceedings of 64 pages were published in February 1950. At the price of 50 cents per copy for members of the two organizations and 75 cents to others it was hoped that the project would pay for itself. The production costs ($275.00) were estimated correctly but we underestimated the time required in the Secretary's office for handling the project and did not allow for unpaid orders. About 75 copies now remain and it is hoped that their sale will further reduce the deficit.

However, as one of the direct consequences of the undertaking, the Society has enrolled 15 new members from among the Entomologists, more than double the number previously listed in that biological field.

EXPERIMENTAL DESIGNS

(1) A SURVEY OF TYPES OF EXPERIMENTAL DESIGNS

GERTRUDE M. COX

Institute of Statistics, University of North Carolina

Abstract

PLANNING a piece of research in any field involves a certain order of procedure. This will in general have three parts (1) a statement of the objectives, (2) a description of the experiment covering such points as the selection of experimental treatments, decision regarding accuracy of measurements, selecting the experimental units, determining the general condition under which the test shall be made and specifying the experimental design and (3) an outline of the method of the analysis of the results.

Methods for increasing the accuracy of experiments may be classified into three types: (1) increasing the size of the experiment, (2) refining the technique and (3) handling the experimental material so that the effects of variability are reduced. This may be done by careful selection of the material, by taking additional measurements that provide information about the material or by skillful grouping of the experimental units into an efficient plan.

Using the ideas of confounding and grouping of the experimental units as the criteria, the various types of experimental designs have been classified. The following types of designs were presented, illustrated and discussed: (1) Complete block designs including completely randomized, randomized blocks, latin square and cross-over designs.

TABLE 1—ANALYSES OF FIELD EXPERIMENTS, 1942–1948.
NORTH CAROLINA AGRICULTURAL EXPERIMENT STATION.

	Number of Analyses										
	Soy-beans	Corn	Pas-ture	Pea-nuts	Pota-toes	Cot-ton	To-bacco	Small Grain	Others	Total No.	Per-cent
Completely randomized				22			25		12	59	0.9
Randomized blocks	210	419	665	502	260	362	637	298	558	3911	61.9
Latin Square		23		11	48				9	91	1.4
Split-Plot	122	142	388	50	36	108	70	212	117	1245	19.7
Simple lattice	48	42		54	7		39	43	16	249	4.0
Triple lattice	39	179		5	6	49	5	21	2	306	4.9
Quadruple lattice							59			59	0.9
Balanced lattice	6	45	79	16	3	19	39	19	3	229	3.6
Rectangular lattice						23				23	.4
Lattice square		122		7		2		14		145	2.3
Total	425	972	1132	667	360	563	874	607	717	6317	100.0

(2) Incomplete block designs as balanced incomplete block, balanced lattice, incomplete latin square and lattice square designs. Some partially balanced designs were presented with special emphasis being given to the split-plot designs.

It is impossible to give many general rules which will be helpful in selecting designs. Each experimental situation presents its limitations to be considered. A good working rule is to use the simplest design that meets the needs of the experiment. This is not to say that the more complex designs will be used only rarely. Table 1 gives a summary of the extent to which the different types of designs were used during 1942–1948 at the North Carolina Agricultural Experiment Station. It is noted that almost 62 per cent of the 6,317 analyses were from field experiments arranged in randomized blocks.

Table 2 gives a summary of the relative efficiencies of incomplete block as compared with randomized complete block designs for 1011 analyses of soybean, corn, pasture, peanuts, potatoes, cotton, tobacco and small grain experiments conducted from 1942 to 1948 at the North Carolina Experiment Stations. There is an over all pooled gain in efficiency of 23%. This means that *four* replications of an incomplete block design, on an average, are about as accurate as *five* replications of a randomized block design. This means more efficient use of experimental land and labor.

TABLE 2—RELATIVE EFFICIENCIES OF LATTICE DESIGNS AS COMPARED TO RANDOMIZED COMPLETE BLOCK DESIGNS.

k =		3	4	5	6	7	8	9	10	11	Total No.
Simple Lattice	No.	68	92	42	23	15	5		4		249
	%	110	115	109	115	119	129		127		
Triple Lattice	No.	15	38	60	131	32	14	13	2		306
	%	110	119	129	127	127	128	139	104		
Quadruple Lattice	No.	7		52							59
	%	131		111							
Balance Lattice	No.	87	103	38					1		229
	%	111	128	114					131		
Lattice Square	No.		10	69		54		11		1	145
	%		145	142		139		139		154	

$t = 7 \times 8$

Rectangular Lattice	No.	23
	%	136

EXPERIMENTAL DESIGNS

(2) MULTIVARIATE EXPERIMENTATION

M. H. Quenouille

Marischal College, Aberdeen, Scotland

INTRODUCTION

The use of several variables in experimentation is by no means uncommon. Thus, in field trials the weight of grain and straw may both be measured; in growth experiments, different measurements of animals may be taken over a series of weeks; or in infection trials, the factors believed to influence resistance may be observed. In each of these examples, several concomitant variables are measured and used in the subsequent analysis of the experiment; each contributing in some part to the final estimates and conclusions.

The employment of several variables in this manner presents problems of interpretation, design, analysis and significance, many of which are still unsolved. It is the purpose of the following note to review and extend the existing approaches to, and methods of, such multivariate experimentation.

DEPENDENT AND INDEPENDENT VARIABLES

The first step in the design or analysis of multivariate experiments is to decide which variables are to be regarded as dependent and which as independent. The independent variables may then be employed to reduce the unaccountable variation in the dependent variables using the analysis of covariance technique. The analysis of variance of the dependent variables, thus corrected, may then be used jointly to derive the linear function of them which is most sensitive to treatment differences. This is done by discriminant analysis. For example, the yield of straw may be employed to reduce the variability in grain yield if the straw yield reflects fertility differences but not treatment differences; or alternatively, if straw yield reflects treatment differences, a combination of straw yield with grain yield might be used to test treatment differences.

The decision whether to employ any particular variable as dependent or independent or, in fact, at all, rests partly upon the extent to which each variable reflects treatment differences and partly upon the questions that the analysis is intended to answer. Obviously, if

we are interested in testing only the effect of treatment on individual observations, such as grain weight, then we use only one dependent variable at a time; but, if we desire to determine the manner in which the treatments act upon the observations, we might have several dependent variables. For example, in testing the difference between two diets, we may take several measurements, such as weight, height, and leg-length, and combine them to find the most sensitive index of dietary differences of the type tested.

The choice of independent variables will naturally depend upon what is to be tested, but, if the main aim is to reduce the unaccountable variability, and to detect treatment differences the choice of independent variables will be restricted.* The covariance method will be applicable only if the dependent variable is more sensitive to treatment differences than the independent variable, and the residuals of both variables are correlated. Even when this is so, the method will be inappropriate if the correlation between the residuals is less than that between the residuals prior to the elimination of the treatment effects. For example, the following analysis was calculated for the observed weight increases, y, and food consumptions, x, of chickens receiving four different diets:

	Degrees of Freedom	Sum of squares of y	Sum of products of x and y	Sum of squares of x
Treatments	3	12650	15059	20589
Residuals	32	20510	22273	77349
Treatments + residuals	35	33160	37332	97938
Treatments Variance ratio, F.		6.58		2.84

Correlation between "residuals" = 0.559.
Correlation between "treatments + residuals" = 0.655.

In this example since the weight increases appear to be more sensitive than food consumptions to differences in the diets, and the residuals are highly correlated, it might be expected that the elimination of the

*The need for caution here is emphasized by Mr. M. Healy in the discussion that follows. However, I disagree with Mr. Healy's remark that an independent variable must be altogether independent of treatment effects. Thus it is a common procedure to adjust the measurements in growth experiments for initial weight. If, however, a situation arose in which the only available weights were taken one day after the commencement of the experiment, it would still be possible to adjust using these observations. The question then is where to draw the line. Naturally, great caution must be exercised in choosing such independent variates, and adjustment for weights taken one week after the commencement of the experiment may be very misleading.

effects of food consumption from weight increases would give a more sensitive index of dietary differences. However, since this correlation is less than that observed before the elimination of treatment effects, this expectation is not realised and the covariance analysis gives us:

	Degrees of Freedom	Sum of squares of y	Mean square	Variance ratio F.
Treatments	3	4834	1611	3.54
Residuals	31	14096	454.7	
Treatments + residuals	34	18930		

Evidently the variable should not be treated as an independent variable in this analysis. If, however, the sum of products for treatments had been 5059 instead of 15059, the correlation would have been reduced to 0.480 and the variance ratio would have become 8.38; x could then have been treated as an independent variable.

Thus it is seen that, when it has been decided precisely what has to be tested, it is still necessary to decide whether particular variables should be treated as dependent or independent. This may be done by tests similar to that indicated above.

THE FORM OF ANALYSIS

When it has been decided which variables are to be regarded as dependent and which as independent then the analysis may be carried out in two stages. First, the sums of squares and products for the "residual" and "residual + treatments" should be used in a covariance analysis to eliminate the effects of the independent variables. The degrees of freedom will correspondingly be reduced by the number of variables eliminated. Secondly, the sums of squares and products, thus corrected, may be used in a discriminant analysis as described by Fisher (1948). The coefficients of the "most-sensitive" linear function may be evaluated exactly using the method of divided differences or, by trial and error, minimising the ratio of the residual sum of squares to the residual + treatment sum of squares. Alternatively, we may calculate the coefficients by successive approximation. This approach is useful since it allows the effect of extra variables to be examined and a possible method will now be described.

Suppose the residual or "within" sum of squares and products are given by W_{ij} and the residual + treatment or "total" sum of squares are given by T_{ij} . Let $W_{ij}/T_{ij} = R_{ij}$ and suppose $R_{11} \leq R_{22} \leq R_{33}$ etc., then x_1 is the most sensitive individual variable, but in general

$x_1' = x_1 + \sum_{i=2}^{i} a_i x_i$, where $a_i = 2(R_{11} - R_{i1}) \times \text{sign } T_{i1}$, will be more sensitive. Thus, x_1 can be replaced by x_1' and a second approximation calculated. Two points should be noted:

(a) The convergence will be more rapid, if the units of $x_1 x_2$ are chosen to make the *sum of squares of* x_1 *and sums of products* $x_1 x_2$, $x_1 x_3$, $\cdots$ *comparable in magnitude*;

(b) The condition of the last section for the covariance method to lead to a more sensitive test of treatments is equivalent to

$$R_{ii} R_{jj} - R_{ij}^2 > 0$$

The following examples will serve to demonstrate this method.

Examples

1) The analysis of the last section may be extended as follows:

	Sy^2	Sxy	Sx^2
W_{ij}	20510	22273	77349
T_{ij}	33160	37332	97938
R_{ij}	0.618516	0.596620	0.789775

$a_2 = 2(0.618516 - 0.596620) = 0.044$
$y' = y + 0.044x$

	Sy'^2	Sxy'
W_{ij}	22619.77	25676.36
T_{ij}	36634.82	41641.27
R_{ij}	0.61744	0.61661

$a_2 = 2(0.61744 - 0.61661) = 0.0017$
$y'' = y + 0.0457x$

	Sy''^2	Sxy''
W_{ij}	22707.29	25807.85
T_{ij}	36776.69	41807.77
R_{ij}	0.61744	0.61730

$a_2 = 2(0.61744 - 0.61730) = 0.0003$
$y''' = y + 0.0460x$

The changes in R_{11} are very small and evidently further approximation would not change the discriminant function appreciably.

2) As a second example, weight increases over a longer period were analysed for this experiment. There were three variables to be considered: y, the weight increases; x_1, the food consumed during the last two weeks; x_2, the food consumed during the first three weeks. The sums of squares and products of these variables were

	Sy^2	Sx_1y	Sx_2y
W_{ij}	69224	78148	9953
T_{ij}	107292	122119	32447
R_{ij}	0.64519	0.63993	0.30675
		Sx_1^2	Sx_1x_2
W_{ij}		189031	58545
T_{ij}		253270	93670
R_{ij}		0.74636	0.62501
			Sx_2^2
W_{ij}			89626
T_{ij}			117605
R_{ij}			0.76209

Since $R_{ii}R_{jj} - R_{ij}^2 > 0$ for all i and j, the three variables could be considered as dependent. The analysis then proceeds as previously:

$$a_2 = 0.01, \qquad a_3 = 0.68$$

$$y' = y + 0.7x$$

	Sy'^2	Sx_1y'	Sx_2y'
W_{ij}	127075	119130	72691
T_{ij}	210344	187688	114770
R_{ij}	0.60413	0.63472	0.63336

$$a_2 = -0.06, \qquad a_3 = -0.06$$

$$y'' = y - 0.06x_1 + 0.64x_2$$

	Sy''^2	Sx_1y''	Sx_2y''
W_{ij}	105481	104275	63801
T_{ij}	176059	166872	102094
R_{ij}	0.59912	0.62488	0.62492

$a_2 = -0.05,$ $a_3 = -0.05$

$$y''' = y - 0.11x_1 + 0.59x_2$$

	Sy'''^2
W_{ij}	89663
T_{ij}	150558
R_{ij}	0.59554

The calculation may be stopped when the desired degree of accuracy is obtained or, at any stage, convergence may be accelerated by examining the changes in the coefficients. The exact discriminator, in this example, should be $y - 0.19x_1 + 0.49x_2$ and for this linear function, $R_{11} = 0.59245$.

TESTS OF SIGNIFICANCE

It has already been indicated above that the degrees of freedom of W_{ij} and T_{ij} should be reduced by the number of independent variables eliminated. Where a discriminant analysis is carried out the effect is more difficult to estimate, but may be approximately gauged by decreasing the degrees of freedom of W_{ij} , but not T_{ij} by one less than the number of dependent variables. Thus, if T_{ij} and W_{ij} have n and $n - p$ degrees of freedom respectively and q dependent variables are used, the revised degrees of freedom for R_{ij} will be $n - p - q + 1$.* Alternatively, Bartlett's (1938) test may be used and $-\{n - \frac{1}{2}(p + q + 1)\} \log_e R_{11}$ may be tested as χ^2 with $p + q - 1$ degrees of freedom. This test will usually be less accurate than the above test.

*This will be an exact test for either p or q equal to one.

If we carry out these tests for the above examples, we get the following analysis for example 1:

	Degrees of Freedom	Sy^2	Mean Square	Variance ratio
$T_{11} - W_{11}$	4	14070	3517.5	4.80
W_{11}	31	22707	732.5	
T_{11}	35	36777		

$-\{35 - \frac{1}{2}(3 + 2 + 1)\} \log_e 0.61744 = 15.43$ is a χ^2 with 4 degrees of freedom.

Both of these tests indicate that P lies between 0.01 and 0.001, but it is apparent that there is no increase in accuracy from the inclusion of x. This can be tested using Bartlett's test, since χ^2 with 3 degrees of freedom prior to the inclusion of x is approximately $-\{35 - \frac{1}{2}(3 + 1 + 1)\} \log_e 0.61852 = 15.61$. The difference, which is negative owing to the approximate nature of the test, indicates that no improvement has resulted from the inclusion of x. The improvement in accuracy might also be tested using the original analysis of variance. The appropriate residual sum of squares for this is $33160 \times 0.61744 = 20474$, so that the analysis is as follows:

	Degrees of freedom	Sum of squares	Mean square
Original treatments	3	12650	4217
Improvements due to including x	1	36	36
Treatments	4	12686	
Residuals	31	20474	660
Total	35	33160	

The improvement is, as before, insignificant. For example 2, the corresponding analysis is:

	Degrees of freedom	Sum of squares	Mean square	Variance ratio
Original treatments	3	38068	12689	
Improvement due to including x_1 and x_2	2	5327	2664	1.25
Treatments	5	43395	8679	4.07
Residuals	30	63897	2130	
	35	107292		

$$\chi^2_{(5)} = -\{35 - \tfrac{1}{2}(3 + 3 + 1)\} \log_e 0.59554$$
$$= 16.33$$
$$= -\{35 - \tfrac{1}{2}(3 + 1 + 1)\} \log_e 0.64519$$
$$= 14.24$$

Improvement due to including x_1 and x_2 , $\chi^2_{(2)}$, $= 2.09$.

The two forms of analysis are again in substantial agreement; the value of P being just below 0.01 and the improvement due to including x_1 and x_2 being barely above expectation.

The form of analysis indicated in the previous section is in consequence very useful for rapid investigations of the variables relevant to any enquiry. For this purpose the degree of approximation that is required may be easily gauged. Thus, in the final approximation of example 2, R_{11} is obviously subject to an error of not more than 0.01 and the treatments sum of squares is therefore at most 1073 too small (actually 332). This difference is not of any importance, but, if it had been, a closer approximation could have been taken.

EXPERIMENTS WITH FIXED TREATMENTS

When the treatments are fixed in their application throughout all phases of the experiment, as in a growth experiment, the results may be analysed jointly or separately. Thus, for example, the total growth may be analysed or alternatively, the effect of the treatments on the growth curve may be investigated, depending on which aspect is regarded as important. Bartlett (1947) has given an example of the multivariate analysis of estimated parameters of a growth curve and,

similarly, an analysis may be carried out on the direct observations as in the following example.

Example.

Forty pairs of littermates (mice) were used in an experiment to investigate the effect of a dietary supplement. The weight increases were observed two, four and six weeks after weaning and the differences between pairs, x_1 , x_2 , x_3 were calculated for each period. The sums, sums of squares and sums of products were as follows:

	$\Sigma x_i x_j$			
$i \backslash j$	1	2	3	Σx_i
1	501.20	174.80	84.52	54.8
2		243.40	287.08	26.2
3			742.68	5.0

The mean difference was largest after two weeks although it was not obvious that the subsequent differences could be completely ascribed to this large initial effect. A discriminant analysis might therefore be carried out to test whether the second and third observations give further evidence of treatment effects not contained in the first observation. This may be most easily executed by calculating the regression of a dummy variate on $x_1 x_2 x_3$, and consequently, the discriminant coefficients from the equations.

$$\begin{aligned} 501.20a_1 + 174.80a_2 + 84.52a_3 &= 54.8 \\ 174.80a_1 + 243.40a_2 + 287.08a_3 &= 26.2 \\ 84.52a_1 + 287.08a_2 + 742.68a_3 &= 5.0 \end{aligned}$$

Using successive eliminations, we get

$$\begin{aligned} 182.436a_2 + 257.603a_3 &= 7.088 \\ 257.603a_2 + 728.427a_3 &= -4.242 \\ 364.688a_3 &= -14.249 \\ a_3 &= -0.03907 \\ a_2 &= 0.09402 \\ a_1 &= 0.08314 \end{aligned}$$

The significance of the further observations may then be tested as follows:

	Degrees of freedom	Sum of squares	Mean Square
Treatments (period 1)	1	$\frac{(54.8)^2}{501.20} = 5.992$	
Treatments (period 2, period 1 effects eliminated)	1	$\frac{(7.088)^2}{182.436} = 0.275$	
Treatments (period 3, period 1 and 2 effects eliminated)	1	$\frac{(-14.249)^2}{364.688} = 0.557$	
Residual	37	33.176	0.897
Total	40	40.000	

There is apparently little to be gained by using the additional observations from the second and third periods.

REVERSAL EXPERIMENTS AND EXPERIMENTS WITH ROTATING FACTORS

The reversal experiment is usually analysed by comparing quadratic or cubic components of the two treatment groups. For example, if $A_1B_2A_3B_4$ and $B_1A_2B_3A_4$ are the two treatment groups, $-A_1 + 3B_2 - 3A_3 + B_4$ and $-B_1 + 3A_2 - 3B_3 + A_4$ are compared. By this means it is hoped to eliminate trend and to deal only with observations which are sensitive to treatment differences. However whether such linear functions are the most sensitive will depend upon the form of trend and the stability of the variance from one phase to the next. For many purposes, it will not matter whether the function used is the most sensitive, but it is possible to carry out a more extensive analysis using discriminant functions. This is most conveniently carried out by the regression method as above, but in the following example the method of successive approximation has been used, since this may generally be applied to experiments with rotating factors.

Example.

The weight increases, x_1 , x_2 , x_3 , were observed for forty-eight chick-

ens on a reversal experiment. These were used to calculate $y_1 = x_3 - 2x_2 + x_1$, $y_2 = x_2 - x_1$, $y_3 = x_1$ which were then used in a discriminant analysis as follows:

	Sy_1^2	Sy_1y_2	Sy_1y_3
W_{ij}	48205	−18315	2242
T_{ij}	53124	−20857	2726
R_{ij}	0.90741	0.87812	0.82245
		Sy_2^2	Sy_2y_3
W_{ij}		9912	−2060
T_{ij}		11226	−2310
R_{ij}		0.88295	0.89177
			Sy_3^2
W_{ij}			3018
T_{ij}			3066
R_{ij}			0.98434

The quantity y_2 is obviously more sensitive than y_1 and inspection of the values of R_{12} and R_{23} indicates that little is to be gained from the inclusion of y_1 and y_3 in the discriminant function. We may however test this by calculating the first steps of the analysis.

$$a_1 = -0.01, \qquad a_3 = 0.02$$

$$y_2' = y_2 - 0.01y_1 + 0.02y_3$$

	$Sy_2'^2$	$Sy_2'y_1$	$Sy_2'y_3$
W_{ij}	10201.03	−18752.21	−2022.06
T_{ij}	11556.19	−21333.72	−2275.94
R_{ij}	0.88273	0.87899	0.88845

$$a_1 = -0.01, \qquad a_3 = 0.01.$$

Since these values do not give rise to greatly changed values of R_{ij}, we might try

$$y_2'' = y_2 - 0.03y_1 + 0.06y_3$$

	$Sy_2''^2$	$Sy_2''y_1$	$Sy_2''y_3$
W_{ij}	10809.88	−19626.63	−1946.18
T_{ij}	12249.26	−22287.16	−2207.82
R_{ij}	0.88249	0.88062	0.88149

The inclusion of y_1 and y_3 leaves the value of R_{22} virtually unaltered. As previously this could be tested using the analysis of variance:

	Degrees of freedom	Sum of squares	Mean square
Treatments (y_2)	1	1314	
Treatments (y_1 and y_3)	2	5	2.5
Total Treatments	3	1319	
Residual	44	9907	225
Total	47	11226	

Of more interest in this case is whether the inclusion of y_2 and y_3 , after y_1 has been used, increases the discrimination significantly. To test this, we get the analysis:

	Degrees of freedom	Sum of squares	Mean square
Treatments (y_1)	1	4919	
Treatments (y_2 and y_3)	2	1324	662
Total treatments	3	6243	
Residual	44	46881	1065
Total	47	53124	

Evidently y_2 and y_3 do not cause a significant reduction in the residual sum of squares, and y_1 would be sufficiently sensitive for most purposes.

The above example also demonstrates a possible approach to the

general rotation experiment. Thus, if the three cycles, $A, B, C, A \cdots B, C, A, B \cdots C, A, B, C, \cdots$ are being used, analyses may be carried out testing the treatment differences in each phase and these may then be combined to give a joint test of treatment differences. As previously, where the means and variances are reasonably stable, we may derive little more from this approach than from the usual least-squares analysis.

RESIDUAL EFFECTS

The experiments with rotating treatments considered in the last section do not allow any residual effects to be estimated. For an experiment with three treatments, the residual effects of A, B and C cannot be distinguished from the main effects of B, C and A. Thus, if the residual effects are likely to be of importance an extended design must be used in which several cycles are used. For example, for three treatments, the cycles A, B, C, and A, C, B might be used in all phases so that the first residuals can be differentiated from the main effects. If A_B , A_C , etc., represent the total of treatments A, which follow treatments B, C, etc., then the main effect of A may be estimated by

$$(A_B + A_C - B_C - C_B)/3$$

and the residual effect of A by

$$(B_A + C_A - B_C - C_B)/3.$$

If, therefore we carry out analyses at each phase and use these to discriminate between A_B , A_C , B_A , B_C , C_A and C_B , the resulting discrimination function can be used in conjunction with the above formula to estimate the main and residual effects.

However, the use of such a function assumes that the main and residual effects are influenced in a similar manner as the experiment proceeds. This is very unlikely and it would seem preferable to calculate discriminants for the main and residual effects separately. This may be done by calculating the components due to main effects (residual effects eliminated), and residual effects (main effects eliminated) and using these in a discriminant analysis. For this purpose, since the component due to residual effects (main effects eliminated) is dummy during the first phase, this should be used as an independent variable.

The analysis suggested above is only one of several possible approaches to the analysis of long-term experiments, but, where many years are spent collecting experimental data, it does not seem unreasonable that extra time should be spent on its analysis.

Finally, it must be noted that the above applications of multi-

variate analysis by no means exhaust its possibilities in the analysis of long-term experiments. Where several materials or crops are involved in a cycle, different treatments or treatment cycles may be compared for each crop separately or may be combined in an overall analysis to determine and test the manner in which the treatments act. Such analyses may appear to be extravagant, but if they throw further light upon the treatment behaviour their use will be amply justified.

REFERENCES

Bartlett, M. S. (1938) "Further aspects of the theory of multiple regression." *Camb. Phil. Soc. 36*, 33–40.

Bartlett, M. S. (1947) "Multivariate analysis". *J. R. S. S.* (*B*) *9*, 176–197.

Fisher, R. A. (1948) *Statistical Methods for Research Workers*, Oliver and Boyd, Edinburgh, 10th Edition.

DISCUSSION ON EXPERIMENTAL DESIGN

(1) A SURVEY OF TYPES OF EXPERIMENTAL DESIGNS

GERTRUDE M. COX

A. Hald. My comments will treat only a very small part of the ground covered by Professor Cox and they will be of a more theoretical nature.

Experimental results are obscured by the effect of two types of disturbing factors: (1) the effect of factors displaying *an unknown systematic variation* such as the fertility of the soil, and (2) the effect of factors *varying at random* as for instance errors of measurements. The total experimental error is composed of variations originating from these two sources.

A fundamental problem in most experiments is, therefore, to eliminate as far as possible the effect of systematic factors from the experimental error. Two means are available for this purpose: the *design* of the experiment, i.e., the lay-out of the plots and the allocation of the treatments to them, and the *statistical technique* used in the analysis of the results.

The most commonly used designs such as the randomised block, the latin square, etc. are based on a combination of systematic and random arrangements of the treatments. In this manner part of the fertility variation is eliminated from the experimental error by means of the systematic structure of the experiment *and* the analysis of variance. The remaining variation in fertility is, however, included in the experimental error, and due to randomisation it has become a random variation. Here the "analysis of variance" is taken in its usual sense as the application of the method of least squares under the assumption that the *fertility variation may be estimated or described by a step-function*, which, for example, in the randomised block experiment takes on a new value for every block, but is constant within a block. This method is often unsatisfactory for a large number of treatments because the plots within a block are not even approximately of the same fertility. This fact has led to new designs containing smaller blocks such as the incomplete block designs.

Another possibility for solving this difficulty does not seem to have been fully explored, namely that of eliminating the fertility variation by using another statistical technique, for instance, regression analysis. It would be interesting to know how much could be gained in efficiency by describing the fertility variation in a randomised block experiment by an adequately chosen function. Naturally this depends partly on the fertility variation itself, the gain being largest in cases with large fertility variation. The additional computational labour required by using, for instance, a polynomial instead of a step-function will be considerable.

If it can be shown, however, that it is possible to eliminate nearly all the fertility variation by a regression analysis it follows that *the arrangement of the treatments within the blocks is immaterial*. We may then give up the randomisation and choose an arrangement such that the computations will be relatively easy. In 1929 Neyman proposed using a polynomial for describing the fertility variation in experiments, *where the treatments are arranged in the same order in every block* and where the blocks are placed in succession. In a paper in 1948 I developed a technique using orthogonal polynomials for the same purpose. In about 20 actual field experiments where I have applied this method it seemed to work satisfactorily. The fertility variation

was described by polynomials of up to the fifth degree and the residual variation seemed to be random. In the same paper I explored yet another statistical technique and tried to describe the fertility trend by a simple moving average. The trend was then eliminated by subtraction and a suitably modified analysis of variance of the residuals was used to estimate the treatment means and the experimental error. This method also seemed to work quite well and required much less arithmetic than the regression analysis. What little practical experience I have leads me to hope that this last method may prove to be a valuable one.

The next step must obviously be to compare the results of the various techniques suggested in a large number of widely different practical situations.

H. Åstrand. In the twenties we used systematic plot arrangements with moving averages such as were mentioned by Professor Hald. Following Professor Fisher's visit the latin square became the standard design, and in about 20,000 field trials carried out in the thirties perhaps 80% were latin squares. Randomized blocks are now the most common because they are more convenient for mechanical work and harvesting. More complex designs are used in breeding work, especially in testing a great number of strains. For general research we prefer very simple designs that the farmer himself can carry through. Errors of measurement are small compared with the soil variation not eliminated by the design. The correlation between residual variance and mean yield in 1,200 latin squares carried out in south Sweden was 0.45. Interactions are also much larger on poor soils. Insofar as designs must differ on soils of different fertility, it is not correct to weight the experiments inversely as their variances because this underrepresents the poor soils in the average. We now study trends of soil fertility by soil analyses and try to use simple systematic designs on such prestudied soils.

M. P. Schützenberger. Dans certains types d'expériences les observations sont d'un type non paramètriques—en psychologie par exemple. Les schémas correspondant aux "incomplete block designs" peuvent être avantageusement utilisés pour la comparaison des stimuli car s'ils permettent comme la méthode des comparaisons par paire de tester les interactions ils ont en outre l'avantage d'être aisément utilisables.

R. A. Fisher. The method of fitting polynomials in one or two dimensions is not so new or unexplored as Dr. Hald suggests. For many years Professor van Uven in Holland was exploring these methods: and in the period 1912–1925, before I understood fully the principle of randomisation, I frequently tested, in discussion with my friend "Student" (W. S. Gossett), what could be done with field trials by these means. Of course, the principal difficulty we encountered, once the labour of such analysis had been overcome, was that fitted polynomials in two dimensions might easily absorb so many as 20 or 30 degrees of freedom without removing a corresponding proportion of the residual sum of squares.

Other participants in the discussion included G. Rasch and F. Bernstein.

(2) MULTIVARIATE EXPERIMENTATION

M. H. Quenouille

M. Healy. The question of choice of independent variate needs further discussion. A useful independent variate must adequately reduce the error variance; but

it must also be, not only less sensitive, but altogether independent of treatment effects if misleading results are not to follow. On this point, the analysis of variance by itself is by no means a reliable guide. For example in the case discussed it appears possible that the various diets differ in palatability. Correcting the weight increases for differences in food consumption may therefore lead to results different from those in which the experimenter is interested.

The iterative approach to the best discriminant will be of great use in avoiding lengthy computation aimed at attaining the top of a flat maximum. In practice, however its importance can be exaggerated. In the experiment quoted, the experimenter requires an answer to the questions, "Do these diets give different weight increases and if so, by how much?" Even the first is not adequately answered by stating that the treatments affect significantly a linear compound of weight increase and food consumption while the second part is left unanswered altogether. If a linear compound is meaningful, the experimenter himself is often in the best position to specify the weights.

M. S. Bartlett. While we would probably all agree with the cautions expressed by Mr. Healy, there are many situations for which discriminant analysis is required and it is then important to use the simplest efficient function. There is no point in using complicated functions, as has sometimes been done, if these are not superior to simpler ones. Here the methods described by Mr. Quenouille for examining the effect of introducing further variables should be of value.

Frank Yates. Summing up, the chairman stated that the papers and discussion indicated the importance of the subject and the progress that had been made in the last thirty years. He was a little disappointed that there had been no discussion of the problems arising in long-term agronomic experiments and animal experiments with alternating treatments. He commended these problems to the attention of biometricians.

From his own experience he would like to emphasize the importance of objectivity in the methods of analysis. This was one of the most important and fundamental contributions to the subject made by Professor Fisher. Once the experimenter—or the biometrician—permitted himself the liberty of selection from among a number of alternative methods of analysis, the danger of influencing the results in the desired direction arose. Simplicity in the analysis was also of the greatest practical importance.

Rothamsted experience had indicated the extreme importance of factorial designs. Recent developments in this direction, in particular the use of a single replication or a half or third replication (fractional replication) with the estimation of error from high order interactions, and also split-plot confounding, have proved of the greatest value in providing designs for the investigation of an increased number of factors in experiments with relatively few experimental units. He would commend factorial designs to the attention of experimenters working on animals and human beings who had so far failed to take advantage of the possibilities of this type of design.

Whether in agricultural field trials a large number of simple experiments was better or worse than a small number of more elaborate experiments appeared to have no single answer—it was a balance between advantages and disadvantages. However, a great part of the cost of an experiment on an ordinary farm was unaffected by the size of the experiment and it was therefore often worthwhile to have experiments sufficiently large to include a number of factors.

Other participants in the discussion included G. Rasch and F. Bernstein.

BIOMETRICAL ASPECTS OF BIOLOGICAL ASSAY

(1) BIOLOGICAL ASSAYS WITH SPECIAL REFERENCE TO BIOLOGICAL STANDARDS*

J. O. Irwin

Medical Research Council's Statistical Research Unit
London School of Hygiene and Tropical Medicine

Abstract

The paper contains seven sections. Section I gives a historical account of the development of biological standards from the time of the Budapest International Congress of 1894, when Roux and his colleagues reported on the efficacy of diphtheria antitoxin, until the London W.H.O. Conference of 1949. It includes Hartley's Table of International Biological Standards. There is a brief reference to the history of Statistical Methods; advances in the design of tests are the most noteworthy advance since 1937.

Section II discusses general ideas. If we are given a standard there is no difficulty in defining a unit. The unit is defined as the specific biological activity of a given amount of the standard. It cannot be defined as the given amount itself because we may want to assay against the standard substances which exhibit "the specific activity" but are not necessarily in the same chemical form. "Specific activity" although somewhat tendentious is an unavoidable phrase. It has as its background a working hypothesis which often has to be abandoned as more is learnt about the drug. A substance which initially has been regarded as though it were a pure chemical compound has later often been found to be a mixture of several. The ideal thing is then to enable each of these to be assayed separately, either by biological or

*This paper is published in full in the Journal of Hygiene Vol. 48, 1950.

preferably by physical or chemical means. When the constitution of each is known and they can be synthesised we are approaching the stage when the standard will be unnecessary. To make it unnecessary should be the ultimate aim of research. There is no difficulty in defining potency provided we are prepared to admit that it *may* vary at different levels of dosage or in tests with different species of animals. When this happens the definition is deprived of much of its practical utility, but the results are an indication that more fundamental research is required, until the situation is cleared up.

Section III discusses Statistical Technique. There are subsections on the equivalence of various formulae for fiducial limits, on a new method for the combination of fiducial limits from individual assays to obtain fiducial limits of a pooled result, and on the χ^2 test in Probit Technique when the numbers are small. The true χ^2 distribution is obtained for the particular case of one animal on each of eleven equally spaced doses $(-2\sigma(0.4\sigma)+2\sigma)$, the S.D. (σ) of tolerances on the dosage scale being supposed known. This is found to be fitted satisfactorily by a Pearson Type VI curve. Both very large and very small values of χ^2 occur more frequently than in the 11 d.f. distribution based on normal theory. A general method is given for obtaining all the semi-invariants of the χ^2 distribution when the true mortalities are known. In most cases where the numbers of animals on each dose are small it would appear that a "normal theory" χ^2 distribution with a modified exponent is a satisfactory approximation to the distribution.

Section IV outlines a method for applying probit-technique when the test can be arranged so that one member of each of a number of litters can be put on each dose. Section V discusses shortly the relative value of probits, logits and the angular transformation. Section VI compares a number of miscellaneous rapid methods for the quantal case, of which Kärber's method is on the whole the most useful.

Section VII, the concluding section, gives an account of the international co-operative test of the new standard for Vitamin D.

The paper ends by emphasising how great is the debt which is due to the pioneers who succeeded in getting standards established, people like Dale, Gautier, Gaddum, Hartley and Trevan, thereby enabling many of the new discoveries of medicine to be utilised on a comparable basis throughout the world to the immense advantage of many thousands of sufferers; also to the splendid contribution made to this end by the Biometricians and Statisticians, too many of whom were at the Conference to make it anything but invidious to mention them by name.

BIOMETRICAL ASPECTS OF BIOLOGICAL ASSAY

(2) THE ANALYSIS OF A COLLABORATIVE ASSAY OF THE THIRD INTERNATIONAL DIGITALIS STANDARD PREPARATION*

W. L. M. Perry, M.D.

National Institute for Medical Research, London, England

Abstract

As many of you are aware, digitalis consists of an extract of the leaves of the plant *Digitalis purpurea*. In this respect it is a comparatively simple biological preparation. Chemically, however, it is exceedingly complex and its exact nature is not yet elucidated. It is known, however, that there are present more than 25 glycosides, some of which, like digitoxin, have been prepared in a pure chemical form, others of which are of incompletely proven constitution. Moreover, two samples of leaf may contain widely different proportions of these 25 or more glycosides and in some samples a few of them may be altogether absent.

It is also known that there are often large variations in the potency ratio of two samples of digitalis when the test animal is changed, or when the conditions of the assay in any one species of test animal are changed. In the first case this can be attributed to a varying sensitivity of the animal species to different glycosides which are present in the two samples in different proportions and, in the second place, to quote one example, the different potency ratios obtained in the past between frog assays read at 2 hours and at 24 hours could be attributed to different proportions of a slowly-acting glycoside which produced its effects only after the end of the 2-hour period.

In addition to these facts, neither the cat nor the frog is human, and neither the guinea-pig nor the pigeon suffers as far as I am aware from auricular fibrillation.

These, then, are the major difficulties and they are indeed formidable. In fact, the assay of such heterogeneous material must always be regarded as a makeshift—as a practical necessity in lieu of something better—and we must always beware of attaching too great a degree of precision or of definition to the final estimate of potency.

When it became necessary to establish a Third International Digitalis Standard some time ago an attempt was made to minimise

*The full report of the results and analysis of the collaborative assay, a summary of which was circulated to members of the Conference, will be published elsewhere.

these difficulties by arranging for the proposed Third International Digitalis Standard to be a blend of a number of separate samples of leaf; a similar blend had been used successfully in making the Second International Standard. In this way it was hoped to obtain a more nearly representative mixture with average relative proportions of all the glycosides. In order to detect species differences all the laboratories collaborating in the assay of the proposed new standard were requested to assay in at least two species of animal. At least one of these two assays was to be carried out by one of three recommended methods which were circulated to the laboratories, for cats, frogs and guinea-pigs respectively. In the recommended methods care was taken to ensure firstly randomisation of the animals with certain restrictions about weight and sex, and secondly, as complete control as possible of known sources of variation, such as speed of injection of dose, method of preparing dilutions, and the temporal changes of sensitivity in a stock of laboratory animals. In this way it was hoped to be able to estimate differences between laboratories using one standard method, as well as differences between the alternative methods of assay.

I would like to consider now the question of assigning a potency value to the proposed Third International Digitalis Standard. The results of all the analyses, when combined, give a final estimate of the weighted mean potency of the proposed Third International Digitalis Standard in terms of the Second International Digitalis Standard of 1.0527, with limits of error ($P = 0.05$) of 1.036 to 1.069. There are however, several other considerations to be taken into account in assigning a definite potency level to the new standard preparation.

Firstly Gold and his co-workers have shown that man is much more nearly a cat than a frog—in respect of his reaction to digitalis. The response used in man was the degree of improvement in selected cases of auricular fibrillation, and if this work is accepted as conclusive the potency ratio in the cat—shown in Table I—to which the potency ratio in the pigeon approximates, would be expected to reflect more exactly the expected potency ratio in man.

Secondly, it would be convenient practically if the change to the new standard could be effected without making any change in the regulations. This could legitimately be recommended on statistical grounds only if the best estimate of the potency ratio, namely the combined weighted mean potency, did not differ significantly from unity. This does not hold in the present assay.

Thirdly, there is no evidence at all to show that the standard preparation of digitalis is completely stable. By this I mean that a gross deterioration of the Second International Digitalis Standard over

TABLE I

RESULTS OF COMBINATION OF ASSAYS DONE SEPARATELY FOR EACH METHOD.

Method of Assay	No. of Assays	Potency	Limits of Error ($P = 0.05$)		Weight	χ^2_M	d.f.	P
			Actual	Percentage				
Frog	12	1.0197	0.977–1.064	95.8–104.4	11092	7.65	11	0.7–0.8
Cat	16	1.0874	1.057–1.117	97.2–102.8	25816	24.99	15	0.05–0.1
G.P.	6	1.0363	0.997–1.077	96.3–103.9	13927	6.00	5	0.3–0.5
Pigeon	7	1.0745	1.037–1.113	96.5–103.6	16539	5.11	6	0.5–0.7
G.P. (Other Methods)	6	1.0330	0.995–1.072	96.3–103.8	14654	19.12	5	0.001–0.01
Dog	1	0.8712	0.775–0.978	88.9–112.3	1740	—	—	—

a period of more than 10 years would almost certainly have become noticeable; a change of the order of 5–10% would equally certainly not be evident. I do not wish to labour this point, which is purely an academic one, since there is equally no proof that such deterioration does occur; moreover the precautions taken in the preparation and storage of standard preparations are designed to prevent it; and indeed it is a necessary and fundamental assumption in using standard preparations to assume absolute stability. Nevertheless it may, perhaps, be of some slight significance that I can find no instance, in the small number of cases of which records are available, of a replacement standard preparation proving to be less potent than its predecessor. This will probably be due, in many cases, to improvements in the methods of preparation of the new standard; but on the other hand it may be due to small reductions in the potency of the old standard which had occurred during the several years of its life. Thus, although we may be able precisely to define the potency ratio between the new standard at the beginning of its life and the old standard at the end of its life, we cannot be sure that the same ratio would apply if the old standard were also in the full flush of youth. I do not advance this as an argument in favour of not using the weighted mean potency as the best estimate of the potency of the new standard; I would, however, use it to illustrate the extreme difficulty of proving that any such estimate is exact; and to argue, therefore that undue elaboration of statistical technique is an unnecessary and unwarranted refinement in such cases.

The final assignment of potency must await the decision of the Expert Committee on Biological Standardisation of the World Health Organisation, but such points as these will probably have some bearing on the question. It is to be hoped, therefore, that today's discussion may throw some interesting light upon them.

DISCUSSION ON BIOMETRICAL ASPECTS OF BIOLOGICAL ASSAY

C. I. Bliss. In agreement with most investigators, both Doctors Irwin and Perry have restricted the term "biological assay" to experiments which would test the null hypothesis that the activity assumed for one preparation, the "unknown", does not differ from that observed for a second preparation, the "standard". One further assumes that within the framework of the experiment the unknown is qualitatively identical with the standard. The latter hypothesis is a great convenience in design and analysis, even though we may know in advance that under other conditions the standard and unknown do in fact differ qualitatively.

Given these requirements, it is clear that an experiment in which the response to a single preparation is measured at several dosage levels is not a biological assay, despite its key importance in assay design. Its purpose is to determine whether the preparation is active and, if active, to define its dosage-response relation. It involves only one major assumption, that the dose reaching the site of action is proportional to that applied and measured by the experimenter.

True biological assays depend upon this and additional assumptions which may be used to group them under three headings with increasingly stringent restrictions.

The first of these are comparative assays (Rasch's term), which are of especial interest in research. In addition to the conditions for a dosage-response curve, a comparative biological assay requires that the response to both the standard and the unknown must be measurable in the same units, that the preparations must be compared within self-contained assays of measurable precision and, in some cases, that the response to comparable doses of standard and unknown must not differ significantly. Comparative assays measure the potency of the unknown relative to the standard under specified conditions and determine whether this potency is qualified by the level of response. As an aid in selecting one from several alternative techniques, the inherent precision of each method should be reported. For this purpose it is convenient to use Gaddum's $\lambda = s/b$ (or $1/b$ in quantal assays), with which one can estimate the number of animals required for a given precision or vice versa. λ is needed quite apart from the fiducial limits at $P = .95$ or .99, used commonly but limited to a specific design.

Analytical biological assays (Finney's term) form the second type, in which the active ingredients are presumably identical. These assays are concerned with biological standardization and may make use of the techniques of quality control. Ideally, they involve two additional conditions that are seldom realized in practice. One is that if there are two or more active ingredients, their relative proportions must be constant in all preparations. The other requires that the relative potency must be entirely independent of the assay method, of the test organism or of the level of response. An analytical biological assay answers the question, "What is the potency of the unknown in units of the standard?"

Analytical biological assays should be designed so that their inherent error can be measured. When this shows the requisite stability in repeated tests, past experience in respect to the slope and the standard deviation can be utilized to minimize the error. One should also determine whether the net inter-assay error is negligible, as it seems to be in the digitalis assay, or whether it may be larger than the intra-assay error, as in the penicillin plate-assay. This cannot be decided *a priori* and it is an important function of collaborative experiments to determine the relative size of these two sources of experimental error. This is possible only when each component

assay provides an estimate both of relative potency and of its internal precision. From this viewpoint each collaborative experiment should follow a carefully specified design, so that in subsequent statistical analyses observed differences in relative potency need not be ascribed to obvious differences in assay techniques.

The third type consists of the pass-or-fail assays. These have the sole objective of determining whether an "unknown" preparation of a given drug passes or fails a prescribed standard of potency. Because of its limited objective only the test of significance is of concern and such assays may differ materially from those requiring an estimation of potency. It is possible, for example, that a single dosage level may be preferred theoretically for the unknown. The assay process rather than the individual test is of primary concern. In addition to most of the assumptions underlying the other assays, consumers' and producers' risks or errors of the first and second kind must be established and a method of inspection specified rather exactly. Although analytical assays could be used for this purpose, a well-designed pass-or-fail assay may be much more efficient. The applicability of these techniques to Pharmacopoeial purposes has yet to be explored in the light of modern developments in inspection sampling.

N. K. Jerne. The first biological substance for which an International Standard was adopted was diphtheria antitoxin, a substance contained in the blood-serum of animals that have been immunized with diphtheria toxin. In 1922 it was decided that 62.8 micrograms of a dried serum-preparation kept at the State Serum Institute at Copenhagen was to be the International Standard Unit of diphtheria antitoxin. The assay of diphtheria antitoxin is fairly simple, and the possibility of expressing the potency of different diphtheria antitoxic sera in International Units has never been seriously questioned.

Nearly all assay methods are based on the following procedure: Diminishing amounts of antitoxic serum are measured into a series of test tubes and brought to constant total volume with saline. A constant amount of toxin is then added to each tube so that in the final series of tubes we have a constant concentration of toxin and a diminishing concentration of antitoxin. After all or part of the toxin has been neutralized by the antitoxin we can measure the remaining toxin concentration by injecting these mixtures into the skin of rabbits (or guinea-pigs). The response to a toxin injection is an inflammatory skin-reaction, the diameter of which is a function of the concentration of toxin injected. By plotting against these responses the log antitoxin dose, we obtain curves which seem to be parallel and reach an asymptote corresponding to the response to the toxin concentration in a tube without antitoxin. This toxin concentration (corresponding to the fixed amount of toxin added to every test tube) can be varied arbitrarily and the assay can thus be carried out at different concentration levels, all of which should give us the same potency evaluation of a serum when tested against the standard.

But this is not what we actually find. At very low concentration levels, the log dose/reaction curves are no longer parallel for different antitoxins, and the distance between the curves is quite different from the distance at higher concentration levels. There may be parallel, steep curves at a high concentration level, and unparallel, flatter curves at a low concentration level, for the same two preparations. In a single routine assay, the slopes of the curves will usually not be determined with a precision great enough to show significant departure from parallelism, but there remains the difference in distance between the curves at different concentration levels. This proves that the active substances in the test serum and the standard serum are not the same.

The logical result of these observations would be that the potency of an unknown

antitoxin preparation cannot be expressed in Standard Units. And as these considerations undoubtedly apply to other sera for which an international Unit has been established, the whole foundation of serological standardization is involved. Yet the measurement of the potency of such sera in Units seems often to work quite satisfactorily in practice. This is because we are usually dealing with very powerful sera for therapeutic use. These can be measured and used in high concentrations, so the differences described here are not detected and not important. But in research experiments, and in all cases where very small concentrations of antitoxin have to be measured, the difficulties become apparent.

In the case of diphtheria antitoxin we have found that the observations can be described by assuming that toxin is neutralized by antitoxin in a reversible equilibrium,

$$2A + T \rightleftarrows A_2T.$$

At a high concentration this process goes almost entirely to the right. At low concentrations a large percentage of the two substances remains free, and to neutralize the free toxin almost completely a surplus of antitoxin must be added. This surplus depends on the value of the equilibrium constant

$$\frac{C^2_A \cdot C_T}{C_{A_2T}} = K$$

which is different for antitoxins from different sera.

At high concentrations the influence of K is negligible, and the "potency" of all sera corresponds to the relative potency that would have been found if they had the same K as the standard serum. But at low concentrations discrepancies will show up, depending on the difference in K between the test serum and the standard serum. These discrepancies are sometimes very large. The potency evaluation at a high concentration may for some sera be 10 times as large as when measured at a low concentration.

The antitoxin contents of an unknown serum can thus be described by two quantities: an estimate of the potency it would have had if its K were the same as the K of the standard serum, and an estimate of the actual K from which the neutralizing power of the serum at all concentration levels can be computed. These two quantities can be estimated from two experiments at different concentration levels, and in research experiments this procedure should certainly be followed.

I have assumed that the determination of potency and of K is independent of the *toxin* used in these assays. If toxins of different quality should yield different estimates of K for the same serum, the matter becomes much more complicated.

In routine assays, where the purpose is to determine the potency of highly concentrated antitoxic sera, there are several reasons for disregarding these complications. It seems quite sufficient to evaluate the potency of such sera in the same simple way as has hitherto been the practice, and to express the estimate of potency in Units.

E. C. Fieller. As an alternative to Dr. Irwin's reconciliation of the various formulae advanced for computing the fiducial limits in biological assay, it would seem better to abandon the formulae, and compute the limits arithmetically, from the familiar quadratic equation. This can be regarded as assigning limits to the root α of a simple equation, the coefficients of which are linear functions of normally distributed observations; a similar method serves to assign limits to the roots of *any* equation with coefficients of that nature.

In connection with Dr. Irwin's method of combining the information supplied by several independent assays, alternative methods have been outlined when the residual variances remain constant from assay to assay. One for use when the slope remains constant extended the method of a previous paper. Of the other two for use when the slope varies, one is based on work by D. V. Lindley, the other on work as yet unpublished by K. D. Tocher.

For judging the inherent reliability of an assay method, there may be practical reasons for preferring the ratio b^2/s^2 to that advocated by Bliss ($\lambda = s/b$). For deciding in practice which of two rival assay methods to adopt, it is essential to consider their comparative costs per response as well as their comparative inherent reliabilities.

J. Tripod. I have here the same feelings as a general practitioner at a meeting of specialists, who is picking up what he needs and selecting new methods according to their precision, simplicity and usefulness rather than for their theoretical basis. As a pharmacologist, I look upon biometrics (1) as an aid to the better planning of experiments, (2) to find the probability of differences and (3) to obtain a sound standardization of specific pharmacological properties. But all of these are tools for enabling me to form a final judgment and never a goal in themselves. Like other tools, the value of a biometric method depends on its precision, simplicity, rapidity and economy. Some methods are relatively easy to apply, but in other cases, we encounter great difficulties either in the computations or in collecting a sufficient number of observations.

Given the same experimental data, we have many ways of estimating a relative potency. In "all or none" responses we found, for example, the following values from the same observations of the relative toxicity of 2 narcotics in mice perorally, in percentages

Graphically		Algebraically	
Trevan	32.8	Behrens	33.8
Gaddum NED	32.9	Van der Waerden	34.1 ± 5.7*
Miller-Tainter	34.0 ± 5.7*	Bliss	33.7 ± 5.6*
Prigge	33.3 ± 5.9*	*P = 0.05	

With the method of Miller-Tainter I found the potency in some minutes, while I needed some hours and even some days without a calculator for the other computations. This example is given to show that we are thankful to the statisticians who give us tools which we can use with more rapidity and more economy, but it is also fair to say that such simplifications are deduced from the more complicated methods.

A second point which is very important for me as a pharmacologist is the unifying and internationalization of statistical symbols. This need may not be apparent in the USA, but I am afraid that the slower adaptation of European biology to biometrics is largely due to the confusion created by various symbols. I am very pleased to have heard that this question is under consideration. The teaching of biometrics would then be easier and more people without a good mathematical background would be attracted by the charm of statistics.

This leads me to a third point. It is very well to apply statistical methods to physiology and pharmacology. But I believe that these sciences are also a means to find new treatments for curing sick people. Hence the statistical treatment of clinical

investigations is at least as important as that of animal experiments. It is unfortunate that many differences between new treatments cannot be analyzed because of heterogenity of the cases, wrong planning of the treatment, insufficient number of observations or even lack of controls. On the other hand, good statistical treatment of a specific effect on animals might give interesting hints for the clinic. Such uses in the clinic can only be promoted by teaching biometrics in medical schools or in post-graduate courses.

Sir Percival Hartley. The question of the stability of biological standards is fundamental to the whole process of biological standardisation and the practice of expressing the potency of biological substances in units. It has engaged the attention of those concerned with the preparation of standards in the inter-war years, and continues to do so. It is difficult to devise experiments which will detect small changes in potency, especially in those assays dependent on the determination of the LD50 dose. Other steps had been taken to ensure, as far as possible, the stability of the standards. Since deterioration is almost certainly associated with chemical change, it was considered that if this could be reduced to a minimum stability should ensue. Accordingly, the standards are absolutely dry preparations, sealed in ampoules which are filled with pure dry nitrogen gas, and kept constantly in the dark and at low temperatures. Other experiments, in which the standards had been sent to India and Australia and back several times or exposed to other adverse conditions had given confidence that the preparation of the standards and their subsequent care, had been effective in preserving their potency.

D. J. Finney Dr. Irwin has described the design used for the collaborative vitamin D assays. In this design, the four solutions were compared by *intra-litter* contrasts, but the regression coefficient was based upon *inter-litter* contrast. The problem arises as to whether the true inter and intra litter regression coefficients of response on dose are necessarily identical. In general they will be, but special circumstances might make them unequal; this type of design, often used for other assay problems, would then give a misleading estimation of potency. In any event, the design has the disadvantage that the regression coefficient is likely to be estimated with relatively much less precision than the mean differences in response.

By modifying the design used in the vitamin D assay, it is possible to alter the scheme of confounding between litters and to obtain an intra-litter estimate of regression without upsetting the estimation of the response differences. For example, with a standard preparation, S, and three test preparations A, B, C, a design for litters of four and three doses of each preparation would be made up of sets of nine litters as follows, the suffixes referring to the dose levels:

I:	S_1	A_1	B_3	C_3	IV:	S_3	A_3	B_1	C_1	VII:	S_2	A_2	B_2	C_2
II:	S_1	A_3	B_1	C_3	V:	S_3	A_1	B_3	C_1	VIII:	S_2	A_2	B_2	C_2
III:	S_1	A_3	B_3	C_1	VI:	S_3	A_1	B_1	C_3	IX:	S_2	A_2	B_2	C_2

Three repetitions of this scheme would need 27 litters, and the experiment would then be comparable with the design for 30 litters actually used.

Martin Leopold. On peut mentionner que dans l'essai biologique de la digitale, le taux du calcium peut jouer un rôle très important. En effet, en 1939, le professeur T. Labarse et le Dr. J. van Heveswinghels de Bruxelles ont montré que, par suite de l'administration intraveineuse de Calcium, la dose lethale de digitale pour le chat diminue considérablement. D'où, dans les essais biologiques, la calcémie pourrait être un facteur statistique à contrôler.

Other participants in the discussion included F. Bernstein and G. Rasch.

CONTRIBUTED PAPERS

(1) OUTLINE OF A MATHEMATICAL THEORY OF PECK RIGHT

Anatol Rapoport

University of Chicago

INTRODUCTION

C. Murchison (1935) has identified a number of "quanta of social phenomena" in the *Gallus domesticus* (the chicken). One of these is the so-called Social Reflex No. 1, observed in the movement toward another member of the species; another is Social Reflex No. 2, fighting a member of the species. He then established methods for measuring these quanta: Reflex No. 1 by the velocity of approach and Reflex No. 2 by willingness to fight and by success in battle. As co-variables, Murchison considers mass, momentum per second and kinetic energy per second in Reflex No. 1 and their relation to success in battle. Another co-variable is "social rank," that is, the peck order position of the bird in its own flock.

N. Collias (1942) points out that among hens "social rank" is established by the result of the first encounter. He then proceeds to analyze the correlations between success in combat with a member of a strange flock and other measurable characteristics, such as size of comb, degree of moulting, and social rank in the home flock.

Thus the task of identifying, measuring, and correlating "social quanta" is under way in the study of events which may well be considered simple social phenomena. In these studies, the unit is taken to be the *individual*, and the observables are his individual characteristics and his relations to other individuals within and outside his society (flock).

We wish to attack the problem from a somewhat different point of view and by a different method. Out unit will be the *society* (a small flock of birds), and the observables will be the elements of structure of that society. Our method will be deductive. We shall make certain postulates concerning the determinants of social structure and by probabilistic analysis draw conclusions about the social structures likely to arise.

In the work of Murchison, Collias, and others, where correlation is sought between observables, it is desirable to reduce unknown and variable factors to a minimum, since these tend to reduce the significance of the correlations. But in our (probabilistic) approach, we shall on the contrary be more interested in *random* events. Ideal situations of this sort (where complete randomness reigns) yield no less significant regularities and laws than "deterministic" situations.

GENERAL CONSIDERATIONS

Peck order or peck right is a binary asymmetric relation between each pair in a finite aggregate of individuals (a society). The relation will be designated by $>$ (or by $<$), so that $A > B$ is read "A pecks B," and, of course, $A < B$ is read "A is pecked by B." The relation is asymmetric—if $A > B$ holds, then $B > A$ does not hold, at least not simultaneously. Two facts about this relation (as observed in bird and other societies) are of interest to the mathematical biologist.

1. The relation is not necessarily transitive, $(A > B).\ (B > C)$ does not always imply $A > C$.

2. The set of relations in a given society may change with time, tending to transform the structure of the society into a simple chain (Murchison, 1935; Allee, 1931), thus,

$$A > B > C \cdots > Z. \tag{1}$$

The first observation points to indeterminate factors in the dynamics of such a society, while the second points to determinate factors.

One could, of course, postulate a deterministic dynamics and still get intransitive relations. For example, the peck right between two individuals may be determined by the relative magnitudes of the values of a function $f(x_a , y_b)$, whose arguments are variables associated with *both* individuals, so that $A \gtrless B$ according to whether $f(x_a , y_b) \gtrless f(x_b , y_a)$. Then we may well have

$$f(x_a , y_b) > f(x_b , y_a);\ f(x_b , y_c) > f(x_c , y_b);\ f(x_c , y_a) > f(x_a , y_c). \tag{2}$$

One could, therefore, begin the logical development of the peck right problem by examining some such function $f(x, y)$. The structure of the society would then be determined by the form of f and by the distribution of the characteristics among its members. To account for changes in structure, one could make f dependent either explicitly or implicitly on time. This approach does not seem attractive since the number of functions one may choose for f is prodigious, and one has no *a priori* reason to prefer one to another.

The probabilistic hypothesis, on the other hand, supposes that the

structure of a society is determined *to a certain extent* by the outcomes of chance events. This means, in the last analysis, that we are taking into account our ignorance of the determining factors in the dynamics of the society, but we are formulating this ignorance mathematically as is done in any probabilistic approach to a problem. It is also possible to combine deterministic with probabilistic factors, where, for example, one assigns to events probabilities which are functions of some known quantities.

There is another reason besides mathematical convenience for introducing chance events into the dynamics of animal sociology. If, as is indicated by the observation above, the later stages of a society are more "organized" than the early stages, one could postulate the initial working of chance and an inherent bias in the situation which causes the structure to "gravitate" to a certain form. Thus, if one spills peas on an uneven surface, the initial distribution of the peas will be nearly random, but eventually they will assume positions along the lines of minimum potential energy.

SOME POSSIBLE ASSUMPTIONS

A natural choice of a chance event is the result of an encounter between two individuals. We shall assume that in any such encounter one will enjoy "victory" and the other suffer "defeat." It will be further assumed that the probabilities of the outcomes of such encounters can be computed on the basis of the knowledge of the history of the individuals and their inherent characteristics. The "structure" of a society, that is the distribution of social ranks, will be supposed to be determined by the outcomes of such encounters.

One can make several different hypotheses about factors influencing the outcomes and about factors resulting from them. The following are examples of such hypotheses, not necessarily consistent with each other.

1. The result of the first encounter between any two individuals is equiprobable.

2. The probabilities of the results of the first encounter are functions of the respective characteristics (independent of time) of the individuals concerned.

3. The probability of each encounter depends on the result of the immediately preceding encounter between the same individuals.

4. The result of each encounter depends on the total history of encounters experienced by the individuals concerned.

5. Relative social rank between two individuals is completely determined by the result of their first encounter.

6. The result of each encounter determines the relative social rank between two individuals, but other encounters may ensue in which the rank may be reversed. The probability of the reversal depends on the difference in the social rank between the individuals.

7. The probability of an encounter taking place at all is a function of the difference in social rank.

Etc.

One could begin with any self-consistent combinations of these assumptions and derive its implications. These implications would be conclusions about the probability distributions of various "types" of societies, and, if the frequency of encounters were also known or assumed, one could (in principle) derive expressions for the changes of these probability distributions with time. One should thus obtain a "wave" of probability distribution. The abscissa of this wave would be a particular type of structure of a society (to be defined below); the ordinate would be the probability frequency of this structure; and the whole wave would be moving into the third (time) dimension, changing its shape as it moved, so that after a long time it would exhibit a sharp crest at the type of structure toward which societies tend to gravitate. Thus a surface would be generated by this probability wave which could be considered as a complete representation of the "statistical dynamics" of the society.

TYPES OF STRUCTURE

It remains to define the "type of structure" or simply the "structure" of a society. Evidently a definition is desired which would enable us to recognize structure by observation. In a society of N individuals, each will have $N - 1$ peck right relations of which r will be dominant (the right to peck) and $N - 1 - r$ submissive (the necessity to be pecked). There will then be a distribution of numbers $(r_1 , r_2 , \cdots r_N)$ among the N individuals such that

$$\sum_{i=1}^{N} r_i = \tfrac{1}{2}N(N - 1). \tag{3}$$

This set of numbers $(r_1 , r_2 , \cdots r_N)$ together with all its permutations will be defined as a *structure* of the society. A structure can also be represented by a diagram, where each individual is placed on a level determined by the number r_i of dominant peck right relations he enjoys. Thus a society of four individuals can have exactly four structures, diagrammatically represented in Figure 1. Arrows indicate peck right; the numbers indicate the dominant peck right relations of each individual.

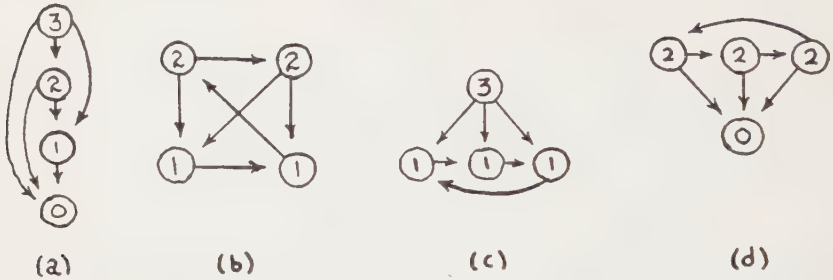

FIGURE 1.

The problem of determining the dynamics of such a society, given a set of assumptions about the occurrence and results of encounters, reduces the problem to the calculation of the probability of occurrence of each structure at each time in the history of the society.

We shall first make the following simple assumptions.

1. The results of the first encounter between any two individuals are equiprobable.

2. The result of the first encounter determines permanently the peck right between the individuals concerned (i.e., the winner will always peck the loser).

Thus the shape of the probability distribution wave is permanently established (independent of time) as soon as all the $\frac{1}{2}N(N-1)$ encounters have taken place. An alternative interpretation is that we are limiting our observation to the period when the structure depends on the results of first encounters only.

THREE INDIVIDUALS

Since only one type of structure is possible for a society of two individuals, namely, (1, 0), i.e., the "dominant" individual pecks one individual, and the submissive one pecks none, we shall begin with the case of three individuals.

There are $2^3 = 8$ possible sets of outcomes of the three encounters. Each outcome leads to one of two possible structures, (2, 1, 0) and (1, 1, 1). That is to say, in one type of society, (2, 1, 0), there is one individual who pecks two others, one individual who pecks one other, and one who pecks none. In the other type (1, 1, 1) each individual pecks one other (and, of course, is pecked by one). The respective probabilities of these structures are $\frac{3}{4}$ and $\frac{1}{4}$, since six of the eight outcomes map on the first (simple chain), and two on the second (simple cycle). This solves the problem of the three individuals.

SOME ASPECTS OF THE N-INDIVIDUALS PROBLEM

One would like to generalize the solution to N individuals. However, difficulties are encountered even under the extremely simple assumptions made above.

The probability of an event is defined as the ratio of the number of ways the event can happen as a result of certain circumstances to the possible occurrences, assumed equally likely. Now in a society of N individuals there will eventually take place $\frac{1}{2}N(N-1)$ first encounters. Since there are two possible results to each encounter, the number of all possible occurrences will be $2^{\frac{1}{2}N(N-1)}$. To compute the probability of a certain structure, one must calculate how many of those "sets of results" will map on a particular structure, that is a particular distribution of peck rights $(r_1, r_2, \cdots r_N)$ or any permutation thereof. This would be a simple matter if there were a one-to-one correspondence between each set of results and each *ordered* set $(r_1, r_2, \cdots r_N)$. Then the probability of the structure $(r_1, r_2 \cdots r_N)$ would be simply the number of *distinguishable* permutations of the set divided by $2^{\frac{1}{2}N(N-1)}$. That the mapping is not in general one-to-one is shown by the following counterexample.

Each set of results can be represented by a skew-symmetric matrix, for example, in a society of four individuals by

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{pmatrix} 0 & -1 & -1 & 1 \\ 1 & 0 & 1 & -1 \\ 1 & -1 & 0 & -1 \\ -1 & 1 & 1 & 0 \end{pmatrix} \end{array} \tag{4}$$

where a 1 in the X-row, Y-column indicates $X > Y$ and a -1, $X < Y$. Note that the matrix (4) maps on the structure (1, 2, 1, 2). But so does the matrix

$$\begin{array}{c} \\ A \\ B \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{pmatrix} 0 & -1 & 1 & -1 \\ 1 & 0 & -1 & 1 \\ -1 & 1 & 0 & -1 \\ 1 & -1 & 1 & 0 \end{pmatrix} \end{array} \tag{5}$$

Without solving the problem of determining the probability distribution of structures for all N, it is nevertheless possible to re-word it in such a way as to state the results in terms of properties of such skew-symmetric matrices. We shall here state the result proved elsewhere (Rapoport, 1949a).

Consider the class $\mathfrak{M}$ of all skew-symmetric matrices (a_{ij}) of order N, all of whose non-diagonal elements are units, that is, $a_{ij} = 1$ or -1; $a_{ij} = -a_{ji}$. We then have the following

Theorem 1. Let (a_{ij}) be a matrix of class $\mathfrak{M}$ and let

$$S_k = \left\{ \sum_{j=1}^{N} a_{ij} \, , \, \sum_{j=1}^{N}{}' \, a_{2j} \cdots \sum_{j=1}^{N} a_{Nj} \right\}$$

be the ordered set of its row sums. Let p_k be the number of distinguishable permutations of the set S_k and m_k the number of matrices in $\mathfrak{M}$ giving rise to the set of row sums S_k. Then the probability of the structure $(r_1 \, , \, r_2 \, , \cdots r_N)$ where

$$r_i = \frac{1}{2}\left(\sum_{j=1}^{N} a_{ij} + N - 1 \right) \tag{6}$$

is given by

$$P(r_1 \, , \, r_2 \, , \cdots r_N) = 2^{-\frac{1}{2}(N)(N-1)} p_k m_k \tag{7}$$

All the structures $(r_1 \, , \, r_2 \, , \cdots r_N)$ are such that the r_i may be put in the form (6).

A general method of finding m_k for any S_k and any N is at this time unknown to the author. However, for N not too large, say, $N \leqq 6$, the probabilities of all structures can be easily computed by a systematized method (Rapoport 1949a).

A sample table for $N = 5$ is given below.

Structure	Probability
(4, 3, 2, 1, 0)	15/128
(3, 3, 3, 1, 0)	5/128
(3, 3, 2, 2, 0)	15/128
(3, 3, 2, 1, 1)	30/128
(4, 2, 2, 2, 0)	5/128
(4, 2, 2, 1, 1)	15/128
(4, 3, 1, 1, 1)	5/128
(3, 2, 2, 2, 1)	35/128
(2, 2, 2, 2, 2)	3/128

The foregoing results implicitly predict the distribution of structures in a large number of societies of a given number of individuals, provided the hypotheses used in their derivations are valid. On the other hand there are strong indications that in any actual flocks significant bias factors are at work. We should not therefore expect the distribution in actual flocks to follow the computed distributions given here. They may be expected to hold figuratively speaking only *in vacuo*. However a liaison between the above theoretical considerations and experimental procedure can be accomplished in a variety of ways.

1. The above results can be used as a theoretical stepping stone for complicating the theory by the introduction of biases known to be important.

2. A crude prediction can be made even on the basis of the results obtained, namely, that in a situation where peck right is determined by the results of the first encounter, the probability distribution will depart from that predicted here so as to weigh more heavily those structures which are closer to complete hierarchy (simple chain). For example, on the basis of random results of encounters, the probability of a complete hierarchy in a society of three individuals is $\frac{3}{4}$. With the introduction of some bias, we would expect it to exceed that value.

3. Whereas the statistical work of correlating physical characteristic and social rank with success in combat is most significant when the opponents are *not* evenly matched, the results of the present theory can be best realized when they are matched as evenly as possible. Thus experimental work can be extended to another range.

"SOCIAL MUTATIONS"

A society of N individuals may have $n(N)$ possible structures. Let us now suppose that from time to time encounters take place between pairs of individuals. The result of each encounter may be either the preservation of the old peck right relation between the two individuals or its reversal. If the peck right relation is preserved, then certainly the structure of the society is also preserved. On the other hand, if the peck right relation is reversed, the structure may be changed, or it may not. For example, the structure S_1 : (2, 1, 0) of the three-individual society is diagrammatically represented as in Fig. 2(a)

A reversal of peck right between individuals 2 and 0 changes the structure to S_2 : (1, 1, 1), as in Fig. 2(b)

But the reversal of peck right between 2 and 1 results again in the structure (2, 1, 0) with the individuals simply relabelled as in Fig. 2(c).

We shall refer to changes in structure as "mutations" and will

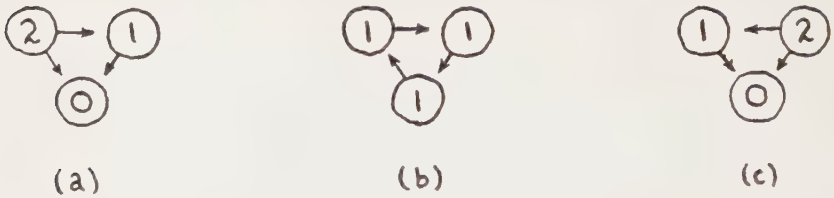

FIGURE 2.

denote the mutations $S_i \rightarrow S_j$ by S_{ij} . The probability of the occurrence of S_{ij} will be denoted by a_{ij} .

In general, not every mutation S_{ij} can be accomplished by a single reversal of peck right. When the mutation $S_i \rightarrow S_j$ cannot be accomplished by a simple reversal, the probability of S_{ij} is 0 ($a_{ij} = 0$). Otherwise a_{ij} ($i \neq j$) will depend on the probability of an encounter between a pair of individuals which may result in S_{ij} and on the probability of the victory going to the previously "submissive" individual. However, in estimating the ultimate fate of the society (the limiting probability distribution of its possible structures), one may use the parameters a_{ij} directly and compute the limiting distribution in terms of these. The computation of the a_{ij} in terms of the probabilities of encounters and reversals is a separate problem. We shall first state the problem of determining the limiting distribution in terms of the a_{ij} .

Let $S_i(t)$ be the probability of occurrence of the structure S_i at the time t. Take as a unit of time the average interval between encounters. Then

$$\begin{aligned} S_1(t) &= a_{11}S_1(t-1) + a_{21}S_2(t-1) + \cdots + a_{n1}S_n(t-1) \\ S_2(t) &= a_{12}S_1(t-1) + a_{22}S_2(t-1) + \cdots + a_{n2}S_n(t-1) \\ S_n(t) &= a_{1n}S_1(t-1) + \cdots + a_{nn}S_n(t-1) \end{aligned} \tag{8}$$

Note that the a_{ii} are the "identity mutations," that is, a_{ii} is the probability of the *preservation* of the structure S_i . We denote the matrix of the probabilities by (a_{ij}). The attention of the reader is called to the fact that the subscript notation is the reverse of the conventional one, a_{ij} being the element of the i-th *column* and the j-th *row*. We have chosen this departure from convention in order to keep the suggestion that a_{ij} stands for the probability of S_i mutating to S_j .

If $S(t)$ is the vector $\{S_1(t), S_2(t) \cdots S_n(t)\}$, equation (8) may be written in vector-matrix notation thus,

$$S(t) = (a_{ij})S(t-1) \tag{9}$$

Hence by iteration,

$$S(t) = (a_{ij})^t S(0) \tag{10}$$

and

$$S(\infty) = \operatorname*{Lim}_{t\to\infty} (a_{ij})^t S(0) \tag{11}$$

Theorem 2. There exists a matrix $(\alpha_{ij}) = \operatorname{Lim}_{t\to\infty} (a_{ij})^t$, such that the columns of (α_{ij}) are all identical. The vector represented by any of these columns is the limiting distribution vector $S(\infty)$, and is independent of the initial distribution $S(0)$.

This result is known to statisticians familiar with the Markoff process. A proof based on mathematical induction and multiplication of matrices is given in another paper of the author (Rapoport 1949b).

CALCULATION OF THE LIMITING DISTRIBUTION

The proof of Theorem 2 implies the existence of a vector S invariant under the transformation (a_{ij}), that is,

$$(a_{ij})S = S; \sum_i S_i = 1 \tag{12}$$

The vector S is the limiting distribution vector and can be found by solving the system of $n-1$ linear equations,

$$\begin{aligned} S_1 &= a_{11}S_1 + a_{21}S_2 + \cdots + a_{n1}\left(1 - \sum_{i=1}^{n-1} S_i\right) \\ S_2 &= a_{12}S_1 + a_{22}S_2 + \cdots + a_{n2}\left(1 - \sum_{i=1}^{n-1} S_i\right) \\ S_{n-1} &= a_{1(n-1)}S_1 + \cdots + a_{n(n-1)}\left(1 - \sum_{i=1}^{n-1} S_i\right) \end{aligned} \tag{13}$$

Denote by (b_{ij}) the $n-1$-rowed matrix obtained by deleting n-th row and n-th column of the matrix $(I - (a_{ij} - a_{nj}))$, where I is the identity matrix, and denote by $(b_{ij})^{(i)}$ the matrix obtained by replacing the i-th column of (b_{ij}) by the vector $\{a_{n1}, a_{n2}, a_{n(n-1)}\}$. Then by Cramer's Rule the components of the limiting distribution vector S are given by

$$S_i = \frac{|\,(b_{ij})^{(i)}\,|}{|\,(b_{ij})\,|} \qquad (i = 1, 2, \cdots n) \tag{14}$$

EXAMPLE: SOCIETY OF THREE INDIVIDUALS

In case $N = 3$, $n = 2$, and the system (13) reduces to a single equation,

$$S_1 = a_{11}S_1 + a_{21}(1 - S_1) \tag{15}$$

whose solution is

$$S_1 = a_{21}/(1 - a_{11} + a_{21}) \tag{16}$$

Then obviously

$$S_2 = 1 - S_1 = (1 - a_{11})/(1 - a_{11} + a_{21}) \tag{17}$$

We can now make assumptions concerning probabilities of encounters and victories, and on the basis of these assumptions compute a_{11} and a_{21} . If an encounter between any pair of individuals is equally likely and the probability of victory does not depend on the peck right relation existing before the encounter, that is, it is $\frac{1}{2}$ for each individual, the calculation of the a_{ij} is quite simple. Since the mutation S_{21} can occur as a result of any encounter (Cf Figure 2(b)), provided the peck right relation is reversed, we have $a_{21} = \frac{1}{2}$. On the other hand, S_1 is preserved in 5 out of 6 possible encounters as can be seen from Figure 2, Hence $a_{11} = 5/6$, and

$$S_1 = \frac{1/2}{1 - 5/6 + 1/2} = 3/4; \qquad S_2 = 1/4 \tag{18}$$

Note that this distribution is also the initial distribution for $N = 3$ as has been shown above to result from random initial victories.

The "mutation" method, however, enables us to introduce biases which may depend on inherent properties of individuals and on their social rank. Thus the probability of encounter between individuals of wisely different social rank may be taken to be smaller than that between individuals of nearly equal rank. Likewise the probability of victory of a dominant individual may be taken to be greater than that of the submissive individual, etc. Hysteresis phenomena may likewise be introduced, that is, the dependence of victories on the number of past victories enjoyed by the individual, etc.

LITERATURE

W. C. Allee. *Animal Aggregations, A Study in General Sociology*, University of Chicago, 1931.

N. Collias. "Statistical Analysis of Factors Which Make for Success in Initial Encounters Between Hens," *The American Naturalist, 77*, 519–538, 1943.

C. C. Lienau. "Quantitative Aspects of Organization," *Human Biology, 19*, 163–216, 1947.

C. Murchison. "The Experimental Measurement of a Social Hierarchy in Gallus Domesticus," *Journal of General Psychology, 12*, 3–19, 1935.

A. Rapoport. "Outline of a Probabilistic Approach to Animal Sociology I," *Bulletin of Mathematical Biophysics, 11*, 183–196, 1949a.

A. Rapoport. "Outline of a Probabilistic Approach to Animal Sociology II," *Bulletin of Mathematical Biophysics, 11*, 273–281, 1949b.

A. Rapoport. "Outline of a Probabilistic Approach to Animal Sociology III," *Bulletin of Mathematical Biophysics, 12*, 7–17, 1950.

N. Rashevsky. *Mathematical Theory of Human Relations*, Principia Press, Bloomington, Indiana, 1948.

(2) ON THE MATHEMATICAL THEORY OF THE EQUILIBRIUM OF INTERACTIONS BETWEEN PROTEINS AND OTHER SUBSTANCES

Enzo Boeri

Istituto di Fisiologia, Università di Napoli

A protein P has n sites capable of binding another molecule A, the active concentration of which is x. The interaction between P and A proceeds along n steps of the type

$$PA_{i-1} + A \rightleftharpoons PA_i \qquad i = 1, 2, 3, \cdots n.$$

Each one of these steps is characterized by an equilibrium constant K_i. It has been known since a long time (e.g. Adair 1925) that the saturation y of P with A varies between zero and unity according to a relation of the form

$$y = \sum_{i=1}^{n} i x^i \prod_{j=1}^{i} K_j \Big/ n\left\{1 + \sum_{i=1}^{n} x^i \prod_{j=1}^{i} K_j\right\} \tag{1}$$

It is also known that for the case $n = 1$, the relation (1) is a rectangular hyperbola. Moreover, (1) reduces again to a rectangular hyperbola in cases in which $n > 1$ and the relative values of the successive equilibrium constants are determined solely by statistical factors. In those instances

$$K_{i-1}/K_i = \frac{n - (i-2)}{n - (i-1)} \frac{i}{i-1} \tag{2}$$

see for instance Klotz 1946. In many instances $n > 1$ and the relation K_{i-1}/K_i is more complex than in equation (2). This is the case when the successive binding of A modifies the reactivity of P in a way more complex than it is to be expected by the gradual decrease of free spaces on P. Neighbour ligated A molecules may mutually interact (Pauling 1935), electrostatic factors may be present (Klotz, Walker and Pivan 1946), or eventually the affinity of the active groups on P may be thought as having values distributed in such a way as to be adequately described by a Gauss error function (Pauling, Pressman and Grossberg 1944, Karush and Sonenberg 1949). In these cases the plots of y versus x do not give hyperbolic, but sigmoid curves.

The purpose of the present paper is to call attention on the analogy of the aforesaid problem with that of mono- and multi-layer adsorption. Let now x represent the relative pressure of a gas, the ratio of

actual pressure to the saturation pressure; n is the number of layers and y the saturation of the gas on the adsorbent. Then, according to the BET theory of multilayer adsorption (Brunauer, Emmett and Teller 1938)

$$y = \sum_{i=1}^{n} icx^i \Big/ n\left\{1 + \sum_{i=1}^{n} cx^i\right\} \tag{3}$$

where c is a constant. For $n = 1$ a plot of y versus x gives a rectangular hyperbola (Langmuir's monolayer) and for $n > 1$ sigmoid curves.

In both equations (1) and (3) if we let nf represent the denominator, we have

$$y = x\, df/f\, dx \tag{4}$$

Equations (1) and (3) differ for the fact that (3) has but one constant, whereas (1) has n constants (although all may be derived from K_1 by some relation).

Strangely enough, the analysis of the interaction equilibrium between P (a protein) and A (another substance) is sometimes pretty well represented for $y \leq 0.7$ by the inverse transformation of the BET equation (3) for n layers, of

$$x = \frac{x_s cy[1 - (n+1)y^n + ny^{n+1}]}{n(1-y)[1 + (c-1)y + cy^{n+1}]} \tag{5}$$

the constants x_s and c being graphically determined according to the method of Wolman and Andrews (1948). Also the inverse transformation of Pickett's simplified equation (1945) may be used (see Boeri 1947). It is surprising to see how often n turns out to be the right number, in agreement with the experiment. Examples are presented.

Thus far, I was not able to give sound logical reasons for this coincidence.

REFERENCES

(1) Adair, G. S., *Journ. biol. Chem. 1925, 63,* 529.
(2) Boeri, E., *Enzymol. 1947, 12,* 114.
(3) Brunauer, S. P., Emmett, H., and Teller, E., *Journ. Amer. Chem. Soc. 1938, 60,* 309.
(4) Karush, F. and Sonenberg, M., *ibidem 1949, 71,* 1369.
(5) Klotz, I. M., *Arch. Biochem. 1946, 9,* 109.
(6) Klotz, I. M., Walker, F. M., and Pivan, R. B., *Journ. Amer. Chem. Soc. 1946, 68,* 1486.
(7) Pauling, L. *Proc. Natl. Acad. Sc. US 1935, 21,* 186.
(8) Pauling, L., Pressman, D., and Grossberg, A. L., *Journ. Amer. Chem. Soc. 1944, 66,* 784.
(9) Pickett, G., *ibidem, 1945, 67,* 1955.
(10) Wolman, D. H. and Andrews, L. J., *ibidem, 1948, 70,* 457.

(3) CONTRIBUTION À L'ANALYSE FACTORIELLE D'UN TEMPS D'INCUBATION

Daniel Schwartz, Jean Cuzin, et André Renier

Mosaique du Tabac

Abstract

Le temps d'incubation pour des pieds de tabac inoculés avec le virus de la Mosaique a fait voir une variation suivant la souche du virus, le génotype de la plante et les conditions de milieu. Les auteurs ont étudié certains facteurs pouvant affecter la taille et le taux de croissance de pieds d'une variété fixée de *Nicotiana Tabacum* L. avant leur inoculation avec une suspension homogène de protéine-virus purifiée. Ils inoculèrent, le même jour, la feuille 18 sur chacun de 265 pieds. Cinq jours plus tard, en moyenne, les feuilles inoculées furent amputées et en fin d'expérience, on dénombra 120 pieds mosaiqués et 145 pieds sains.

Une étude statistique fut faite des facteurs suivants: (1) état du pied (= 0 pour pied sain et 1 pour pied malade), (2) longueur de la feuille 18 trois semaines avant l'inoculation, (3) longueur de la feuille 18 le jour de l'inoculation, (4) quantité de feuilles sur le pied le jour de l'inoculation, (5) taux de croissance de la feuille 18 pendant la semaine avant l'inoculation et (6) taux de croissance du pied pendant la semaine avant l'inoculation.

Le coefficient de corrélation partielle de la variable dépendante (1) avec chacunes des autres était $r_{12,3456} = -.27$, $r_{13,2456} = .25$, $r_{14,2356} = -.24$, $r_{15,2346} = -.09$, et $r_{16,2345} = .07$. La limite de signification pour $P = .05$ était de ± 0.12.

Les résultats furent interprétés au moyen de courbes de croissance schématiques des feuilles 18 pour un pied malade et un pied sain. Ainsi, par exemple, parmi des pieds de feuilles 18 égales ou croissant à des taux égaux au jour de l'inoculation, la longueur de la feuille 18 était en moyenne moindre trois semaines plus tôt chez les pieds qui devinrent malades; de sorte qu'ils croissaient relativement plus rapidement avant l'inoculation. Une explication semblerait possible: c'est qu'une substance ayant une concentration proportionelle à l'âge du pied mais non au stage de dévelopement de la feuille pourrait s'être diffusée dans la feuille retardant la propagation du virus.

Une expérience faite plus tard est mentionée brièvement, touchant l'effet de l'illumination. Le fait de placer une feuille à l'obscurité dans la quinzaine précédant l'inoculation raccourcit considérablement le temps d'incubation.

(4) FREE HAND CURVES IN ESTIMATING THE POTENCY OF HUMAN SERA AGAINST TOXOPLASMA

Ch. A. G. Nass

Institute of Preventive Medicine, Leiden

Abstract

Seventy human sera were subjected to a "Sabin-Test", in which constant volumes of a standard culture of the blood parasite Toxoplasma were exposed to each of three doses (1:100, 1:20, and 1:4) of each serum. After staining, a slide count of the number of unstained parasites in 100 gave an estimate (y) of the percentage kill of Toxoplasmas by the given dose of serum. The potency of each serum was to be evaluated from its three y-values corresponding to the coded log-doses $x - 1$, x and $x + 1$.

It was assumed provisionally that the dosage-response curves for the individual sera could be superimposed by shifting them along the log-dose axis, as if the sera were different dilutions of a single active agent in an inert carrier. With a single curve of the expected response (Y) plotted against the log-dose x it would be possible to estimate the content of agent in any given serum from its observed y-values. In the absence of analytical techniques, this assumed curve has been fitted visually. The method is not recommended as a labor-saver.

The three responses, $y(x - 1)$, $y(x)$ and $y(x + 1)$, for each of the 70 sera were listed in order of increasing $\sum y$ and averaged in 14 sets of five. In the first figure the values of $\bar{y}(x - 1)$, $y(x)$ and $\bar{y}(x + 1)$ were plotted against $\bar{y}(x)$ and fitted visually with three curves. $Y(x - 1)$, $Y(x)$ and $Y(x + 1)$. In this first approximation the relative potency of a serum was estimated from the middle dose alone by setting $\bar{y}(x) = Y(x)$, rather than from the more stable estimate $\sum\bar{y} = \sum Y$. In a later stage of fitting, each of the 14 sets was replotted against $Y(x)$ at a level such that $\sum(\bar{y} - Y) = 0$. From the original assumption that all sera were solutions of a single active agent, the diagram was necessarily symmetrical, so that at any point on the $Y(x)$ diagonal the horizontal distance to the $Y(x + 1)$ curve equalled its vertical distance to the $Y(x - 1)$ curve.

Two transformations of the abscissae of these curves aided in adjusting the free-hand diagrams to their final form, the ordinate remaining the same throughout. In the second figure the abscissa was changed to $Y(x - 1) + Y(x) + Y(x + 1)$, so that sets could be located directly by

the condition $\sum\bar{y} = \sum Y$. While the curves were not symmetrical, rectangles drawn parallel to the axes had two opposite corners on the $Y(x)$ curve and the other two corners on the $Y(x - 1)$ and $Y(x + 1)$ curves. Finally, with the aid of the first two graphs to determine the log-dose x of the active agent, the curves were replotted in a third chart against x with $x = 0$ at $Y = 50$. The $Y(x - 1)$, $Y(x)$ and $Y(x + 1)$ curves then became sigmoid and identical except for position, with lower and upper asymptotes at $Y = 0.4$ and 82% respectively. An additional scale on the third figure facilitated reading x for a given serum directly from its observed value of $\sum y = \sum Y$.

The three graphs emphasized different aspects of the data and thus complemented each other. In making the final adjustments, the three curves were held to the rules for transforming one graph to another and were so placed as to minimize the runs of points on one side of a curve. When fitting was completed, the number of runs was nearly exactly what would be expected in a random sequence, as tested statistically.

Of the three degrees of freedom for each serum, one was lost in computing its potency. The variance components for the other two degrees of freedom have been defined and evaluated on the assumption of a binomial distribution of the observed y's about their expected Y's. One, the "slope variance", measured the failure of the curves for the individual sera to follow the general $Y(x)$ curve. This was significantly larger than the second or residual "error variance". It is apparent that all sera could not be represented by one curve or a single active agent.

DISCUSSION ON CONTRIBUTED PAPERS

(1) OUTLINE OF A MATHEMATICAL THEORY OF PECK RIGHT

A. Rapoport

M. Schutzenberger. The paper of Rapoport shows us the broader scope of biometry. It also shows that there is some basic difference between statistical biology and mathematical biology even if we are sure that the laws of a rational science of biology will be those of a mathematical biology. This leads to two points.

The first is the importance of observation and experiment in the selection of initial assumptions. This is especially true in fields more complicated than the example chosen by Rapoport, that of human relations for example. The actual situation, when analyzed with adequate statistical tools, will offer us assumptions that we never would have thought of, a priori.

Secondly, we feel that the real need of mathematical biology is not the mechanical application of the sort of mathematics which was so successful in physics. New mathematical tools, must be acquired, especially devised for biology. We wish to congratulate Dr. Rapoport for doing this by showing how simple combinatorial analysis can be used in biological situations. Although the statistical and the mathematical schools have different points of departure, it is to be hoped that we will meet in the construction of rational models which can be used to describe the structure of social groups, models which possibly can be used to modify such group structure. If we take care always to formulate our models so they are capable of experimental disproof, how can we fail to build a rational biology?

J. B. S. Haldane. Mr. Rapoport trouvera la discussion d'un problème analogue dans le dernier chapitre du premier volume de "The Advanced Theory of Statistics" de Kendall, où il s'agit de la consistance logique d'une échelle de préférences.

(2) ON A MATHEMATICAL THEORY OF THE EQUILIBRIUM OF THE INTERACTIONS BETWEEN PROTEINS AND OTHER SUBSTANCES

E. Boeri

Leopold Martin. La communication du Dr. Boeri est très intéressante. En effet, quoique biochimiste experimental le Dr. Boeri tend à étudier plus intimément le fonctionnement intime de la matière vivante du point de vue physio-chimique. L'isotherme d'adsorption Langmuir considérée entrautre ici à d'ailleurs été largement utilisée par Hinshelwood dans son livre "Chemical Kinetics of the Bacterial Cell" et, dans le cas particulier de l'adaptation du B. Piocyanique à la Streptomicine, par R. Linz et L. Martin (C. R. Soc. Biol. mars 1949).

NEWS AND NOTES

Chicago Meeting

At the Annual Meeting of the American Statistical Association, December 27-29, 1950, in Chicago in the Congress Hotel, The Biometric Society (ENAR) and The Biometrics Section of the American Statistical Association will hold jointly the following sessions:

December 27, 10–12 A. M.

Topic: **Statistical problems in radio-biology.** Chairman: A. E. Brandt. Papers: (1) Gene mutations in populations, Bruce Wallace. (2) Some tracer chemistry experiments with proteins, S. Lee Crump. (3) Metabolism of labeled carbon compounds, Hardin B. Jones.

December 27, 2–4 P. M.

Topic: **Theory of variance components.** Chairman: W. J. Youden. Papers: (1) The present status of variance component analysis, S. Lee Crump. (2) Testing a linear relation among variances, W. G. Cochran. (3) Application to regression and to errors of measurement, John Tukey.

December 27, 4–6 P. M.

Topic: **Measurement of Morbidity**. Chairman: Harold Dorn. Papers: (1) Is the household survey essential in securing morbidity statistics? S. D. Collins and T. D. Woolsey. (2) Prepaid medical care as a source of morbidity data, Neva R. Deardorff. (3) Experience of associated hospital service, Allen Thompson.

December 27, 4–6 P. M.

Topic: **Precision of measurements.** Chairman: W. Edwards Deming. Papers: (1) The specification of precision of measurements, Churchill Eisenhart. (2) The estimation of precision of measurements, Frank E. Grubbs. (3) Estimate of precision of textile instruments, John C. Whitwell.

December 27, 8–10 P. M.

Party

December 28, 9–10 A. M.

Biometrics Section business meeting.

December 28, 10–12 A. M.

Topic: **Statistical methods in pharmacology and immunology.** Chairman: Lloyd Miller.

Papers: (1) Collaborative bio-assays, Lila Knudsen. (2) Statistical methods in immunology, Herbert C. Batson.

December 28, 2–4 P. M.

Topic: **Applications of variance components.** Chairman: G. W. Snedecor.

Papers: (1) Variance components as a tool for the analysis of sample data, Walter A. Hendricks. (2) Consistency of estimates of variance components, R. E. Comstock and H. F. Robinson. (3) Use of components of variance in preparing schedules for the sampling of baled wool, J. M. Cameron.

December 28, 4–6 P. M.

Topic: **Sample survey techniques.** Chairman: W. F. Callander.

Papers: (1) A consumer survey, Arnold J. King. (2) Approaches to agricultural price statistics, F. E. McVay and Henry Tucker. (3) Problems in rural surveys, R. L. Anderson and A. L. Finkner.

December 29, 9–10 A. M.

The Biometric Society business meeting.

December 29, 10–12 A. M.

Topic: **Statistical methods in medicine.** Chairman: Hugo Muench.

Papers: (1) Survival curves for special diseases, Joseph Berkson. (2) Some designs used in clinico-physiological experiments, Donald Mainland. (3) Multivariate analysis in medical research, James A. Rafferty.

December 29, 2–4 P. M.

Contributed papers.

December 29, 4–6 P. M.

Contributed papers.

Cleveland Meeting

A meeting of The Biometric Society, Eastern North American Region, will be held on December 27, 28 and 29 jointly with the meetings of the A.A.A.S. in Cleveland. Several sessions will be devoted to a symposium on mathematical biology and biometry with a number of invited speakers participating.

Members desiring to present 15 minute papers at the meeting are requested to send in the titles of the papers, together with a 200-word abstract, to N. Rashevsky, Committee on Mathematical Biology, The University of Chicago, 5741 Drexel Avenue, Chicago 37, Illinois.

Heterosis Conference

A Heterosis Conference was held June 13 to July 13, 1950 at Iowa State College. The purpose of the conference was to summarize the available information on methods by which the heterotic effects may be attained, the known causes which are responsible for these effects and the problems which require further research for their solution.

As a supplement to the Heterosis Conference, a Methods Workshop was held from July 3 to 14. Statistical methodology for analysis of data from breeding experiments and statistical aspects in the design of such experiments were presented and discussed. John Gowen acted as chairman of the Heterosis Conference. Jay L. Lush, Iowa State College, and R. E. Comstock, Institute of Statistics, were joint leaders of the Workshop.

Thomas Parran, The University of Pittsburgh, Graduate School of Public Health has called to our attention the first bulletin of this Graduate School. The teaching and research activities of the Department of Biostatistics are aimed primarily at the development of methods for the statistical appraisal of the health problems of groups: the community, the family, and special aggregates such as the population in industry and in school. Mr. Parran writes, "The curriculum has been devised to provide a progressive demonstration of the means by which statistical reasoning applied to the several medical and biological sciences can help to solve health problems of groups through the determination of health needs, evaluation of efforts to meet these needs, and the measurement of the influence of social and biological factors on health and disease in man." . . . **Walter D. Foster,** Biometrician, Department of Biochemistry, West Virginia University, received his Ph.D. degree in experimental statistics at the Institute of Statistics, Raleigh, and is now serving as statistician for the Nutritional Status Project in the Northeastern Region. . . . **David J. Finney,** University of Oxford, England, was married to **Mary Elizabeth Connolly** on April 11. . . . **C. I. Bliss,** is recovering from a heart attack. . . . **Marianne Bernstein** is in Oslo, Norway. She writes, "After having been in five different European countries, I must say that Norway is the most austerity ridden country in Europe. So far I have been most impressed by the climate in Italy,

especially in Rome. During the six weeks I was there it only rained three days. I was able to do a lot of sightseeing. Sometimes my father came along explaining; he had studied art in Italy. I endulged in Roman cooking. What a delight to be able to eat twenty different vegetables, numerous kinds of fruits and unrationed meat and butter. I was most impressed in Rome by **G. Gini's** home. Several rooms are used as studies, and the walls are covered with book shelves." . . . **Ralph Bradley** from McGill University, Montreal, has joined the staff as associate professor in the Department of Statistics, Virginia Polytechnic Institute. He is doing research on rank order statistics. Mr. Bradley received his B.A. and M.A. degrees in mathematics from Queen's University and his Ph.D. degree in mathematical statistics at the University of North Carolina. . . . **David Duncan**, senior lecturer in statistical methods at the University of Sydney, Australia, will join the statistical staff of Virginia Polytechnic Institute on September 1. Mr. Duncan holds a B.Sc. degree in Agriculture, a B.S. degree in mathematics from the University of Sydney and a Ph.D. degree in mathematical statistics from Iowa State College. . . . **William Feller,** formerly with the Department of Mathematics, Cornell University, Ithaca, has been appointed Eugene Higgins Professor of Mathematics at Princeton University. . . . **Donald Mainland,** Professor of Anatomy, Dalhousie University, Halifax, Nova Scotia, Canada, since August 1, is Professor of Biostatistics, Department of Preventive Medicine, New York University—Bellevue Medical Center. . . . **Paul L. Munson,** Research Associate, Department of Pharmacology, Yale School of Medicine, has been appointed Assistant Professor of Dental Science, Harvard School of Dental Medicine. . . . **J. E. Morton,** is on leave from Cornell to serve as Chief, Statistical Research and Development Staff, Division of Housing Research, Office of the Administrator, Housing and Home Finance Agency, Washington, D. C. . . **Allan E. Paull,** who has been at the Grain Research Laboratory, Winnipeg, Manitoba, Canada, is now with Abitibi, Power & Paper Company, Toronto. . . . **H. Fairfield Smith,** formerly Statistician with the Rubber Research Institute of Malaya, is now a Professor with the Institute of Statistics, Raleigh, North Carolina.